

THÈSE DE L'UNIVERSITÉ DE LYON

Délivrée par

L'UNIVERSITÉ CLAUDE BERNARD LYON 1

ÉCOLE DOCTORALE INTERDISCIPLINAIRE SCIENCES-SANTÉ

DIPLÔME DE DOCTORAT

(arrêté du 7 août 2006)

soutenu publiquement le 29 avril 2010

par

Dr DUCLOS Antoine Eugène Henri

TITRE :

SÉCURITÉ DU PATIENT EN CHIRURGIE THYROÏDIENNE :

INTÉRÊT DU SUIVI DES COMPLICATIONS PAR CARTES DE CONTRÔLE.

Directeur de thèse : Pr Cyrille COLIN

JURY de thèse : Pr Bernard BURNAND, Rapporteur

Dr Pierre DURIEUX, Examineur

Pr Fabrice MENEGAUX, Rapporteur

Pr Jean-Louis PEIX, Président

Pr Vincent PIRIOU, Examineur

Dédicace

Je tiens à exprimer ma reconnaissance à Cyrille Colin. Depuis 7 ans maintenant, j'ai la chance de collaborer avec vos équipes et de bénéficier de votre connaissance du milieu hospitalier. Vous m'avez donné le goût de l'évaluation en santé. Veillant à mon épanouissement professionnel, vous avez toujours su être à l'écoute et laisser le temps à mes compétences d'éclore librement. À de nombreuses reprises vous m'avez accordé votre confiance. Si je participe aujourd'hui à autant de travaux passionnants, c'est à vous que je le dois.

Je remercie sincèrement Jean-Louis Peix d'avoir joué le jeu de l'évaluation et de la remise en question. Se sentir observé, tel un 'animal de laboratoire', doit être quelque peu déroutant au quotidien. Votre engagement inébranlable dans cette démarche m'impressionne autant que votre capacité à toujours progresser.

Je remercie Pierre Durieux de m'avoir accueilli dans son unité où j'ai compris que la pensée d'un homme ne se jugeait pas à ses titres mais bien à son originalité et son esprit critique. Merci à Joël Ménard également d'avoir su dessiner des itinéraires et m'insuffler la confiance nécessaire pour tracer mon propre chemin.

Je tiens à remercier Bernard Burnand, Fabrice Ménégaux et Vincent Piriou d'avoir accepté, sans hésitation, de participer à ce jury de thèse. La complémentarité de vos domaines d'expertise et la sagesse de vos jugements constituent le gage d'une évaluation complète de ce travail.

Je tiens à témoigner mon amitié sincère à Jean-Christophe Lifante avec qui travailler demeure un plaisir. Jean-Christophe, l'Atlantique ou les péripéties du CHU n'ont pas entravé l'avancée de nos projets. Nous avons encore beaucoup à explorer. A nous de prendre la liberté et de trouver les moyens pour continuer à le faire ensemble. Au passage, *grazie mille* à Pietro Soardo pour sa contribution à nos travaux.

Je tiens également à adresser une dédicace à mon ami Nicolas Voirin, sans qui ma vision de la recherche ne serait pas exactement la même aujourd'hui. Nicolas, échanger avec toi est un enrichissement permanent et je reste convaincu que nous pouvons changer le cours des choses !

Une fois de plus, je suis heureux de remercier Sandrine Couray-Targe pour son inépuisable soutien, sa science du PMSI et à qui je dois beaucoup. Je remercie également Sandrine Touzet de m'avoir initié patiemment à l'évaluation des pratiques cliniques.

Un merci particulier à Philippe Messy et Marion Teisseire pour leur bonne humeur. Merci à Françoise Mony, Caroline Bouvard et Pascale Arlandis qui nous rendent la vie plus facile. Merci à Stéphanie Bourdy et Laetitia Bouveret pour leur collaboration efficace sur CATHY, à Audrey Bestion, Angélique Denis et Anne Termoz sur d'autres projets.

Un clin d'œil amical à Stéphanie Polazzi et Coralie Brault, adeptes des heures supplémentaires, Isabelle Jaisson-Hot, Pascale Guerre et Christell Ganne, équipe de choc du JEM, ainsi qu'à Anne-Marie Schott, Touria Hajri, Laure Guittard et toute la 'bande des Tilleuls'.

Un grand respect pour l'énorme travail réalisé par toutes les équipes du pôle IMER. Merci aussi à tous les internes, externes et étudiants en Master que nous accompagnons au cours de leur formation et qui nous obligent à progresser.

Enfin, au nom de la solidarité entre thésards, j'adresse un message d'espoir à Florence Canoui-Poitrine, Claire Chabloz, Thomas Fassier, Jean-Marie Januel, Marie-Annick Le Pogam...

Et bien sûr, une dédicace très spéciale à mes proches. Une pensée affectueuse pour Axelle, Gatien, Julie, Romain, Solesne, Thomas, William et Yann, restés fidèles malgré la distance. Bravo à ma mère pour son écoute à toute épreuve, à mon père pour son sens du devoir, à Céline (et la troupe) parfois trop loin mais toujours là, à Caroline pour sa générosité un brin décalée et à Jérôme pour nos débats 'Duclosiens'. Et merci à celle qui partage ma vie et mes envies.

À Ève,

*« Jusqu'où l'Homme peut-il aller dans le contrôle de son environnement physique ?
Dans quelles limites ce contrôle dépend-il de l'intelligence humaine, et dans quelle mesure
dépend-il du hasard ? »¹*

Walter A. Shewhart, 1939.

¹ Shewhart WA. *Statistical Method from the Viewpoint of Quality Control*.
Washington DC: Graduate School of the Department of Agriculture, 1939.

SOMMAIRE

1. Préambule	8
2. État des connaissances	11
2.1. La sécurité du patient à l'hôpital	11
2.1.1. Une prise de conscience internationale	11
2.1.2. Le développement récent de la sécurité du patient.....	14
2.1.3. Des indicateurs pour mesurer la sécurité des soins	16
2.1.4. L'analyse systémique des causes d'évènements indésirables	18
2.1.5. Garantir la sécurité du patient au bloc opératoire.....	21
2.2. La carte de contrôle	25
2.2.1. Apport du Dr Shewhart au contrôle qualité dans l'industrie.....	25
2.2.2. Principes d'application d'une carte de contrôle	28
2.2.3. Différents types de cartes applicables en chirurgie	31
2.3. La chirurgie de la thyroïde.....	38
2.3.1. La pathologie thyroïdienne : rappel anatomo-physiologique et indications opératoires.....	38
2.3.2. La thyroïdectomie : bref historique et conduite à tenir chirurgicale	42
2.3.3. Suivi des complications liées à la chirurgie thyroïdienne	44
2.3.4. La mesure de la sécurité du patient en chirurgie thyroïdienne	50
3. Objectifs	52
3.1. Évaluer la faisabilité et l'utilité d'un suivi d'indicateurs par cartes de contrôle dans la pratique courante d'une équipe chirurgicale	52
3.2. Développer une carte de contrôle adaptée au suivi des complications de la chirurgie thyroïdienne	52
3.2.1. Comparaison de performance entre une carte à limites approximées et une carte à limites exactes	52
3.2.2. Comparaison de performance entre une carte à limites brutes et une carte à limites ajustées sur les caractéristiques des patients	52
3.3. Étudier l'effet du neuromonitoring sur la performance individuelle du chirurgien en traçant la courbe d'apprentissage liée à son utilisation	52
4. Approche commune à l'ensemble des travaux	53
4.1. Population étudiée et sources d'informations.....	53
4.1.1. Technique opératoire.....	53
4.1.2. Design d'étude	55
4.1.3. Protocole d'évaluation	56

4.2. Indicateurs d'évaluation	58
4.2.1. Paralysie récurrentielle.....	59
4.2.2. Hypoparathyroïdie secondaire.....	60
4.2.3. Méthodes d'ajustement	61
4.3. Cartes de contrôle développées	62
4.3.1. Cartes de Shewhart.....	62
4.3.2. Cartes CUSUM	64
5. Travaux scientifiques répondant aux objectifs.....	67
5.1. Étude d'applicabilité de la carte de Shewhart au contrôle qualité des résultats de la chirurgie thyroïdienne (<i>article 1</i>)	67
5.2. Travail méthodologique sur le développement d'une carte adaptée au suivi des complications de la chirurgie thyroïdienne (<i>articles 2 & 3</i>).....	73
5.2.1. Performance d'une carte à limites exactes vs. approximées	75
5.2.2. Performance d'une carte à limites ajustées vs. brutes	83
5.3. Étude de l'effet du neuromonitoring sur la performance du chirurgien et courbe d'apprentissage liée à son utilisation par carte CUSUM (<i>article 4</i>).....	89
6. Synthèse.....	97
6.1. Discussion des travaux réalisés	97
6.1.1. Intérêt de la carte de contrôle	97
6.1.2. Limites.....	102
6.2. Perspectives	105
6.2.1. Application des cartes de contrôle aux hôpitaux français	105
6.2.2. Le chirurgien, l'artisan et le pilote	107
6.2.3. Futurs développements liés à la sécurité du patient en chirurgie	112
7. Conclusion.....	117
8. Remerciements	118
9. Bibliographie.....	119
10. Annexes	134
10.1. Lien entre volume d'activité et indication opératoire (<i>article 5</i>).....	134
10.2. Fiche de codage	142

1. Préambule

La sécurité du patient s'impose aujourd'hui comme un élément central des politiques de santé et devient l'affaire de tous à l'hôpital. Afin d'améliorer la qualité de sa pratique et ainsi de minimiser le risque d'évènement indésirable, tout chirurgien doit s'interroger sur sa performance quotidienne et tirer profit de ses erreurs pour en prévenir la récurrence.

Le suivi d'indicateurs représente une méthode de choix pour suivre objectivement l'évolution des complications opératoires au cours du temps. Néanmoins, l'utilisation d'indicateurs basés sur le résultat d'une chirurgie implique que leurs variations reflètent de réelles variations liées à la sécurité de cette chirurgie. Si certaines variations traduisent des modifications significatives de sécurité, justifiant davantage d'investigations et des changements dans la prise en charge chirurgicale, d'autres variations peuvent être inhérentes à la qualité des données collectées en routine ou au recrutement des patients. Une part importante de cette variabilité est également liée au hasard, c'est-à-dire à ce que nous ne connaissons pas ou ce que nous ne savons pas mesurer. Une multitude de facteurs, qui surviennent à différents moments et interagissent les uns avec les autres, influencent ainsi le résultat d'une chirurgie, rendant sa compréhension particulièrement complexe (figure 1).

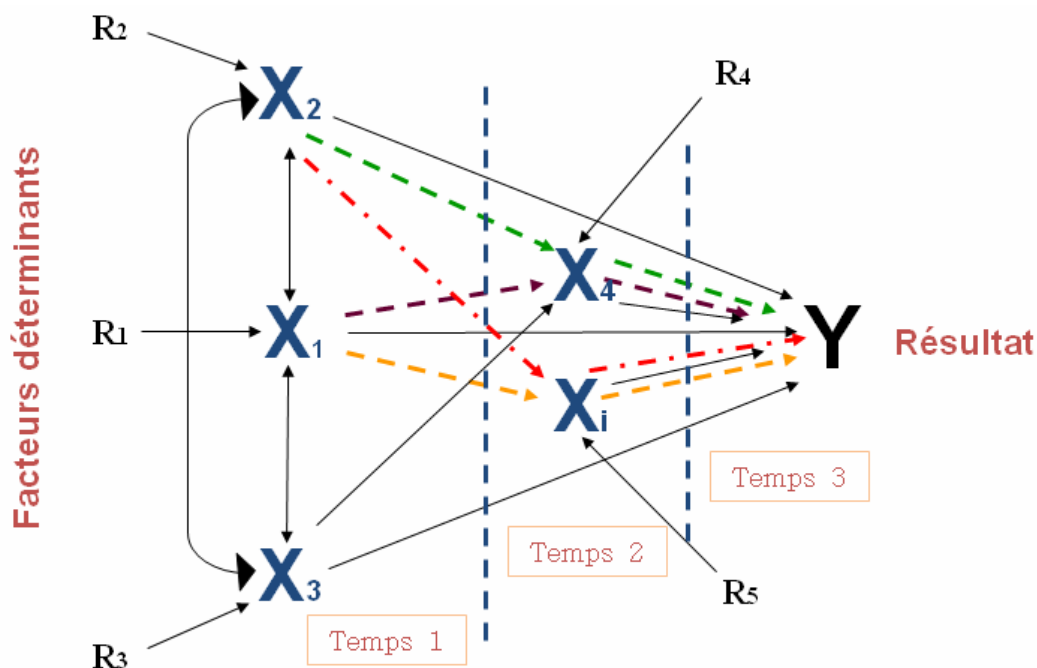


Figure 1. Représentation schématique des facteurs influençant le résultat d'une prise en charge

Au-delà du simple obstacle technique lié à leur production en routine, ces indicateurs posent donc la question de leur validité et de l'interprétation adéquate de leur variabilité. Une réactivité insuffisante ou au contraire excessive à leurs variations peut mener à des décisions inappropriées, conduisant parfois à la réorganisation inopportune des prises en charge chirurgicales. A partir de quel seuil d'alerte sur ces indicateurs convient-il d'engager une action afin de garantir la sécurité du patient opéré ? Et quelle action pourrait s'avérer pertinente pour rétablir ou améliorer la sécurité du patient ? Pour le chirurgien, cette double interrogation peut se résumer à 'quand agir' et 'comment agir', en réponse aux variations des indicateurs dont il dispose ?

La maîtrise statistique des processus apporte une solution à ce problème en donnant du sens aux variations observées. Mise au point au début des années 1930 dans l'industrie pour contrôler la qualité des produits manufacturés, cette méthode tend désormais à s'étendre au champ de la santé. Elle repose sur un outil graphique d'aide à la décision, la carte de contrôle, qui permet de suivre chronologiquement le comportement d'un indicateur afin d'en analyser les variations. En précisant les limites acceptables de ses variations, la carte de contrôle autorise une interprétation plus rigoureuse de la variabilité de l'indicateur, permettant d'aboutir à des décisions plus justes et efficaces basées sur des faits, non sur des impressions.

Pendant quatre années, nous avons suivi la performance collective et individuelle de trois chirurgiens réalisant un nombre élevé d'interventions sur la thyroïde. La mesure de cette performance reposait sur les deux principales complications de la thyroïdectomie : la paralysie récurrentielle et l'hypoparathyroïdie secondaire. Le caractère relativement homogène et standardisé des actes réalisés sur la thyroïde, ainsi que leur volume important comparativement à d'autres types de chirurgie, constituaient un argument facilitateur pour une bonne adaptabilité de la carte de contrôle à la sécurité des soins.

La première partie de ce document rappelle brièvement l'état des connaissances actuelles sur la sécurité du patient à l'hôpital, la carte de contrôle et la chirurgie thyroïdienne. Les objectifs de la thèse sont ensuite détaillés ainsi que la méthodologie communément utilisée pour les atteindre.

La seconde partie du document restitue les travaux scientifiques conduits en cohérence avec chacun des objectifs de la thèse. Nous avons d'abord appliqué une carte de contrôle au suivi des complications de la chirurgie thyroïdienne. Cette expérience de terrain, publiée dans le *British Journal of Surgery* (article 1), a permis de tester initialement l'applicabilité d'un tel

suivi d'indicateurs dans la pratique courante des chirurgiens. Nous avons ensuite mené un travail méthodologique visant à fournir des recommandations pour le développement, la mise en œuvre et l'interprétation d'une carte de contrôle (*article 2*). Une carte de contrôle originale, spécialement adaptée au suivi des complications de la chirurgie thyroïdienne, a été conçue et publiée dans *Quality and Safety in Health Care* (*article 3*). Puis, nous avons expérimenté l'introduction d'une technologie, le neuromonitoring, destinée à guider le geste du chirurgien pour prévenir la survenue de paralysies récurrentielles. Nous avons étudié son influence sur la technique opératoire du chirurgien et tracé la courbe d'apprentissage liée à son utilisation (*article 4*).

La dernière partie est une synthèse discutant les principaux enseignements appris de l'application d'un suivi d'indicateurs par carte de contrôle aux données de la chirurgie thyroïdienne. Les apports et les principales limites méthodologiques de notre travail sont abordés. Une mise en perspective de nos résultats est proposée au regard des approches actuelles, visant à mieux comprendre la performance individuelle du chirurgien et à améliorer le travail en équipe au bloc opératoire.

En annexe, nous présentons une comparaison entre hôpitaux de la qualité de l'indication chirurgicale dans le cancer de la thyroïde selon leurs volumes d'activité respectifs (*article 5*).

2. État des connaissances

2.1. La sécurité du patient à l'hôpital

2.1.1. Une prise de conscience internationale

La communauté médicale a perçu depuis longtemps l'importance des affections iatrogènes, nommées aujourd'hui 'événements indésirables liés aux soins'. C'est en chirurgie aux États-Unis qu'ont été réalisées les premières analyses de morts évitables au début du vingtième siècle.¹ Les réunions de morbi-mortalité, au cours desquelles les complications opératoires étaient analysées de manière qualitative, se sont répandues dans l'exercice quotidien des équipes chirurgicales.² Mais ce sont des travaux récents qui ont permis de quantifier ces accidents directement liés aux soins, sans lien évident avec la raison initiale de la prise en charge ou les caractéristiques du patient. Ces événements ont été qualifiés de graves quand ils entraînaient un décès, un handicap à la sortie ou une prolongation de la durée d'hospitalisation. Habituellement, leur fréquence et leurs causes ont été explorées rétrospectivement à partir d'audits réalisés dans le dossier médical des patients. Dès 1964, une étude révélait que 20% des patients admis dans un service de médecine hospitalo-universitaire présentaient un événement iatrogénique dont 20% étaient sévères.³ En 1981, une autre étude estimait cette proportion d'événements à 36% des patients,⁴ et dix ans plus tard il était rapporté que 64% des arrêts cardiaques survenus à l'hôpital étaient évitables, principalement liés à l'usage inapproprié de médicaments. Puis la Harvard Medical Practice Study, une étude menée sur une population de 30 195 patients hospitalisés dans 51 hôpitaux de l'état de New-York en 1984, a montré que 3,7% d'entre eux présentaient un événement indésirable d'origine iatrogène.⁵ Parmi ces événements, 14% étaient fatals, 48% étaient associés à une intervention chirurgicale⁶ et 69% étaient liés à une erreur et donc potentiellement évitables.⁷

De nombreuses études ont depuis cherché à mesurer la fréquence de survenue des événements indésirables liés aux soins hospitaliers.^{8,9,10,11} On estime aujourd'hui cette incidence médiane à 9,2%, concernant près d'un patient hospitalisé sur dix. Alors que 7,4% de ces événements entraînent le décès du patient, 44% sont évitables et la majorité est liée à une prise en charge chirurgicale.¹² En France, une enquête nationale prospective sur un échantillon de 8 754 séjours hospitaliers a permis d'estimer une densité d'incidence moyenne de 6,6 événements indésirables graves pour 1 000 journées d'hospitalisation (table I).¹³

Table I. Densité d'incidence d'Évènements Indésirables Graves (EIG) observés pendant l'hospitalisation (nombre d'évènements pour mille journées d'hospitalisation)

Type d'activité	EIG			EIG évitables		
	identifiés pendant l'hospitalisation			identifiés pendant l'hospitalisation		
	Nombre	%	[95% IC]	Nombre	%	[95% IC]
Médecine	106	6,2	[4,9 – 7,5]	46	2,4	[1,6 – 3,2]
Chirurgie	149	7,0	[5,7 – 8,3]	49	2,2	[1,5 – 3,0]
Total	255	6,6	[5,7 – 7,5]	95	2,3	[1,8 – 2,9]

Étant donné la nature complexe de la pratique médicale et la multitude de soins prodigués au patient au cours de son hospitalisation, un taux élevé d'erreurs n'est peut-être pas surprenant. Cependant, un taux d'erreur de 1% demeure largement supérieur à ce qui est couramment toléré dans l'industrie, en particulier dans l'aviation ou le nucléaire. Comparativement à ces modèles d'organisations à haut risque, le système de soins accuse un net retard concernant la prise en compte de la sécurité de ses prestations. Une des raisons invoquée au manque de moyens mis en œuvre pour prévenir efficacement la survenue d'évènements indésirables liés aux soins est la sous-estimation par les professionnels hospitaliers de leur fréquence comme de leur gravité. En dépit de leur nombre important, les évènements graves demeurent perçus comme isolés et inhabituels, n'affectant pas la pratique quotidienne de la plupart des personnels soignants. Du fait d'une culture de la perfection inculquée au cours de leur formation initiale et renforcée par leur responsabilité vis-à-vis du patient, les professionnels de santé acceptent mal l'erreur humaine lorsqu'elle survient et tendent à la considérer comme inacceptable, forcément liée à une faute individuelle.¹⁴ Cela peut avoir des conséquences importantes sur leur bien-être et leur motivation, l'événement indésirable entraînant une 'double victimisation', du patient d'une part et du professionnel de santé d'autre part.¹⁵ Pourtant, les évènements indésirables associés aux soins ne sont pas nécessairement synonymes d'erreurs médicales puisque l'on sait que leur survenue résulte de la conjonction de nombreux facteurs individuels ou collectifs liés à l'organisation de la structure de soins.

L'Institut de Médecine Américain a souhaité 'forcer' une prise de conscience collective de ce problème en publiant le rapport '*To Err is Human : building a safer health system*' en 1999.¹⁶ S'appuyant sur les résultats de deux études majeures,^{5,17} ce rapport estimait que 44 000 à 98 000 patients décédaient annuellement aux États-Unis suite à une erreur médicale potentiellement évitable, représentant un coût total pour la société compris entre 17 et 29 milliards de dollars. Une telle estimation plaçait l'erreur médicale devant les accidents

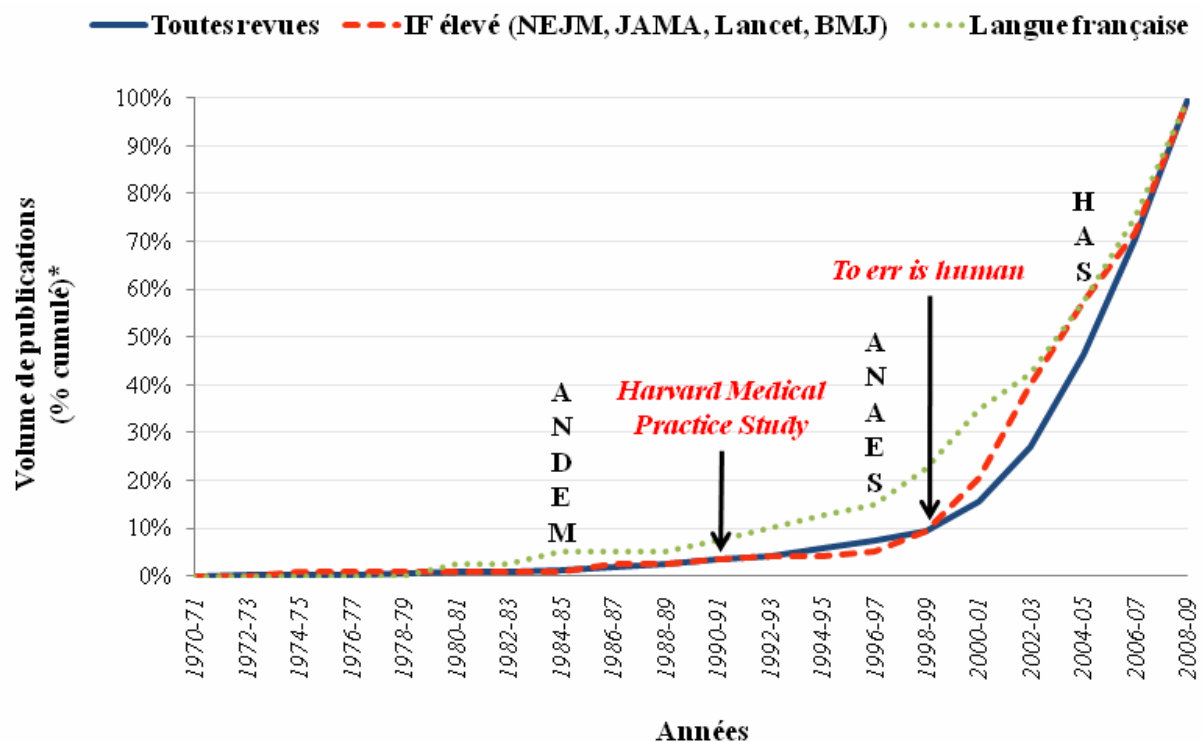
de la route parmi l'ensemble des causes de décès du pays. Considérant le précepte '*primum non nocere*',¹⁸ il n'était pas concevable que le système de soins puisse nuire à la santé du patient. L'erreur médicale, définie comme l'échec d'une action non réalisée selon ce qui était initialement prévu ou comme la planification d'une action inappropriée compte tenu du but à atteindre, était certes humaine mais pouvait être évitée. Afin d'éviter tout amalgame entre événement indésirable et manquement professionnel, une des principales conclusions du rapport était que la majorité de ces erreurs ne résultaient finalement pas de fautes individuelles ou de l'action de groupes de professionnels isolés, mais plutôt de facteurs systémiques liés à l'organisation des soins et à une succession collective de négligences. Par conséquent, la mesure essentielle pour réduire rapidement la survenue de telles erreurs ne reposait pas sur la mise en accusation de personnes mais sur le développement d'une organisation plus sûre, capable de les prévenir ou de les corriger à temps. La diffusion d'une culture de sécurité du patient au sein des hôpitaux était recommandée, comprenant la mise en place d'outils permettant d'en suivre en continu les résultats. Suite à la diffusion de ce rapport, la qualité et la sécurité des soins ont été reconnues comme un problème majeur de santé publique aux États-Unis mais également à travers le monde.¹⁹

2.1.2. Le développement récent de la sécurité du patient

On observe, depuis une vingtaine d'années, un intérêt croissant de la communauté médicale internationale pour l'évaluation et l'amélioration des soins. Après plusieurs décennies où les progrès médicaux se sont focalisés sur les innovations techniques, l'attention se concentre désormais sur les gains de sécurité qui pourraient résulter d'une meilleure prise en compte de l'organisation des prises en charge et des facteurs humains y concourant. La création de structures telles l'*Agency for Health Care Research and Quality* (AHRQ) aux États-Unis ou l'Agence Nationale pour le Développement de l'Évaluation Médicale (ANDEM) en France en atteste, tout comme la très forte croissance du nombre de recherches ciblant la sécurité des soins.^{20,21} En 2004, l'Organisation Mondiale de la Santé (OMS) a lancé une Alliance Mondiale pour la Sécurité du Patient dont les objectifs sont centrés sur l'élaboration d'une politique d'amélioration des soins au sein des états membres. Au plan national, la diffusion d'une 'culture sécurité' parmi les professionnels des hôpitaux apparaît comme un axe prioritaire des actions de la Haute Autorité de Santé (HAS), avec le développement rapide et à large échelle des procédures de certification, de l'évaluation des pratiques professionnelles ou d'indicateurs de qualité des soins.

La recherche sur la sécurité du patient, lorsqu'elle est abordée selon les méthodes de l'épidémiologie quantitative, repose principalement sur deux types d'études : des études expérimentales ou quasi-expérimentales visant à évaluer l'impact d'actions d'amélioration des soins (études d'intervention) et des études observationnelles cherchant à mieux comprendre les facteurs influençant le résultat des soins (succès ou échec d'une prise en charge). Elle s'intègre plus largement dans le champ de la recherche sur les services de santé, qui regroupe l'ensemble des travaux s'intéressant à l'influence du patient, du professionnel de santé ou de facteurs structurels sur la délivrance des soins.²² Il s'agit d'une discipline jeune au regard de la croissance récente du nombre de publications scientifiques s'y référant (figure 2). Elle présente à la fois des chevauchements et de réelles différences épistémologiques avec ce qui est plus communément appelé la qualité des soins.²³ Si les deux termes qualité et sécurité sont inscrits dans un continuum, ils renvoient pourtant à des notions sensiblement différentes. La qualité est étroitement associée à l'idée de référentiel pour le professionnel, construit principalement à partir des données acquises de la science, et les évaluations portent sur le respect de ces référentiels. La sécurité est associée à l'idée de préjudice pour le patient et pointe essentiellement les causes individuelles et systémiques de dysfonctionnement du

système de soins. Dans le cadre de ce travail, nous considérerons que la sécurité du patient se réfère au résultat d'une prise en charge, c'est-à-dire à l'absence de survenue d'un évènement indésirable lié aux soins et dont seule une partie est évitable, par opposition à la qualité des soins, davantage orientée vers l'évaluation des moyens et pratiques mis en œuvre pour éviter la survenue de tels évènements.



* Source PubMed ('Patient Safety')

Figure 2. Évolution temporelle du nombre de publications scientifiques liées à la sécurité du patient
En 2009, 1 241 publications ont été enregistrées dans Medline avec le mot clé 'patient safety', dont 23 dans des revues généralistes à facteur d'impact (IF) élevé et 5 en langue française.

(ANDEM : Agence Nationale pour le Développement de l'Évaluation Médicale ; ANAES : Agence Nationale d'Accréditation et d'Évaluation en Santé ; HAS : Haute Autorité de Santé)

2.1.3. Des indicateurs pour mesurer la sécurité des soins

Partant du postulat que l'évaluation de la sécurité des soins nécessite des données chiffrées pour être rigoureuse, l'utilisation d'indicateurs doit être privilégiée afin d'en suivre l'évolution. Un indicateur est un outil d'aide à la décision, dont l'utilisation répond à un objectif dans un contexte donné. Il n'a d'intérêt que par les choix qu'il aide à faire dans ce cadre. Idéalement, il doit être cliniquement pertinent, fiable (mesure précise et reproductible), validé (mesure reflétant ce qu'il est censé mesurer) et simple d'interprétation.²⁴

Avedis Donabedian a développé une typologie des indicateurs mesurant la qualité et la sécurité des soins en distinguant trois domaines clés : les indicateurs de structure visant à mesurer les ressources mises en œuvre pour la délivrance des soins (comment le soin est-il organisé ?), les indicateurs de processus visant à évaluer les procédures de soins mises en œuvre par les professionnels de santé (que faisons nous ?) et les indicateurs de résultat visant à évaluer les conséquences au décours des soins (quel est le résultat obtenu ?).^{25,26} Classiquement, on considère que les indicateurs de structure et de processus reflètent la qualité du soin, tandis que la mesure de la sécurité des soins se base sur des indicateurs de résultat.²⁷ Une association statistique a été recherchée entre ces trois types d'indicateurs. Il apparaît qu'un lien semble exister entre indicateurs de structure et de résultat, les patients pris en charge par des prestataires de soins à haut volume d'activité présentant un risque moins élevé de complications, principalement dans le cadre de chirurgies complexes en oncologie.^{28,29,30,31} A l'inverse, le lien existant entre indicateurs de processus et de résultat demeure davantage controversé.^{32,33,34,35}

Un indicateur de résultat permet de mesurer objectivement l'état de santé du patient (morbidité ou mortalité) ou bien subjectivement via sa qualité de vie ou sa satisfaction.²⁷ Il reflète d'autant mieux la sécurité du soin que sa mesure est standardisée avec un nombre élevé de patients inclus dans son estimation (minimise le risque de variation aléatoire) et qu'il est ajusté de manière adéquate sur les caractéristiques du patient (*case-mix*) ou du prestataire de soins.^{36,37} En effet, en dépit de son intérêt clinique majeur, un indicateur de résultat ne fournit qu'une estimation indirecte de la sécurité des soins délivrés, ce qui rend son interprétation complexe. Ses variations proviennent schématiquement de la combinaison de 4 éléments : la nature des données disponibles (définition des cas et modalités de recueil), les caractéristiques des patients pris en charge (âge, sexe, comorbidités, etc.), la qualité des soins prodigués (processus de soins) et le hasard (variations aléatoires).^{38,39,40}

Initié aux États-Unis il y a une dizaine d'années,⁴¹ le développement d'indicateurs sur la sécurité du patient (*Patient Safety Indicators*)⁴² extraits directement des bases de données hospitalières,⁴³ représente actuellement une priorité dans l'ensemble des pays industrialisés.^{44,45} Mesurer la sécurité des prises en charge en fonction de leur résultat (taux de mortalité, taux de complications, etc.) doit permettre de comparer la performance de plusieurs entités (pays, régions, établissements) à un instant donné ou bien de suivre la performance d'une même entité au cours du temps. Néanmoins, les tendances observées sur ces indicateurs dépendent de la qualité du codage dans les bases médico-administratives et une validation plus complète des *Patient Safety Indicators* s'impose avant d'envisager leur utilisation routinière (table II).⁴⁶

Table II. Liste des *Patient Safety Indicators* développés par l'AHRQ

N°	Libellé
1.	Complications d'anesthésie
2.	Décès dans des Groupes Homogènes de Malades à faible taux de mortalité
3.	Escarres de décubitus
4.	Échec de manœuvre de réanimation
5.	Corps étranger laissé au cours d'une intervention
6.	Pneumothorax iatrogène
7.	Infections liées aux soins médicaux
8.	Fracture de hanche postopératoire
9.	Hémorragie ou hématome postopératoire
10.	Désordre physiologique et métabolique postopératoire
11.	Détresse respiratoire postopératoire
12.	Embolie pulmonaire ou thrombose veineuse profonde postopératoire
13.	Septicémie postopératoire
14.	Désunion de plaie opératoire en chirurgie abdomino-pelvienne
15.	Difficulté technique pendant l'intervention
16.	Réaction à une transfusion
17.	Traumatisme à la naissance – traumatisme du nouveau-né
18.	Traumatisme obstétrical au cours d'un accouchement par voie basse (avec instrument)
19.	Traumatisme obstétrical au cours d'un accouchement par voie basse (sans instrument)
20.	Traumatisme obstétrical au cours d'une césarienne

2.1.4. L'analyse systémique des causes d'évènements indésirables

Apprendre de ses erreurs pour prévenir leur récurrence constitue une pratique bien implantée dans les industries à haut risque. Depuis longtemps, l'investigation formalisée des accidents y est devenue une priorité pour en identifier les causes de survenue et améliorer la sécurité des services rendus. Dans l'aviation, une attention particulière a été portée aux problèmes de sécurité après la seconde guerre mondiale. En se focalisant sur le facteur humain, les causes d'erreurs mises en avant se sont révélées être davantage liées à des problèmes organisationnels qu'à des fautes individuelles.⁴⁷ Leur résolution a permis une réduction par trois de la proportion de crashes aériens fatals observée en 50 ans, laissant présager de même dans le domaine des soins.⁴⁸ Néanmoins, la reproduction d'une telle démarche à l'hôpital s'est avérée très parcellaire, se confinant aux travaux de rares équipes.⁴⁹

La survenue d'un évènement indésirable lié aux soins résulte typiquement de la conjonction de plusieurs erreurs dont les causes racines proviennent souvent de défauts dans l'organisation du travail.⁵⁰ James Reason a avancé que certains facteurs systémiques pouvaient entraîner des erreurs humaines, tels les sous-effectifs, le recours à des médecins remplaçants ou des infirmières intérimaires, une mauvaise répartition des tâches, une charge de travail ou une fatigue excessives, le manque de communication et de coordination au sein des équipes, l'absence de supervision adéquate des juniors par les seniors, une insuffisance de formation initiale ou continue des professionnels, etc.⁵¹ Sa typologie des erreurs humaines a permis d'en distinguer deux catégories : les erreurs liées à des tâches routinières avec survenue d'un évènement pouvant distraire l'attention et les erreurs liées à une violation consciente des règles de bonne pratique par gain de temps ou absence de contrôle systématique.⁵² Le risque de survenue de telles erreurs est variable et dépend grandement de la complexité de la tâche à accomplir comme des circonstances de la décision, devenant beaucoup plus probable dans une situation imprévue (table III).⁵³ C'est pourquoi, Reason défend un changement en profondeur du système, reposant sur la conception de tâches simples et standardisées facilitant leur réalisation correcte en vue de prévenir d'éventuels oublis ou erreurs d'inattention.⁵⁴

Table III. Exemples de probabilités de survenue d’erreurs selon la tâche réalisée et les circonstances de la décision⁵³

Type d’erreur	Prévalence (%)
Erreur de lecture d’une étiquette	0,3
Erreur ou omission en l’absence d’aide-mémoire	1
Erreur de calcul de dose quand il n’est fait qu’une fois	3
Erreur de reconnaissance d’une alarme	10
Erreur dans des situations de grande urgence ou de grand stress	25

Une autre distinction doit également être considérée entre les erreurs actives, se traduisant par un effet immédiat, et celles latentes qui se révèlent plus tardivement.⁵⁵ Le lien entre l’erreur et l’évènement indésirable n’est donc pas toujours facile à révéler, dès lors qu’il peut être retardé et que la causalité n’est pas nécessairement évidente (Table IV).⁵⁶

Table IV. Exemples d’erreurs avec différents degrés de causalité et d’immédiateté⁵⁶

Type d’erreur	Causalité	Immédiateté
Décès du patient après administration intra-thécale de chlorure de potassium	Forte	Forte
Infection du patient en l’absence de vaccination préalable destinée à la prévenir	Forte	Faible
Décès d’origine coronarienne chez un patient ayant présenté un infarctus du myocarde sans prescription d’aspirine au décours	Faible	Forte
Nouvel épisode coronarien chez un patient ayant présenté un infarctus du myocarde sans prescription de β -bloquants au décours	Faible	Faible

Afin de mieux comprendre les raisons de survenue d’évènements indésirables graves, la mise en place de systèmes de signalement a récemment été décidée dans plusieurs pays industrialisés, selon un mode volontaire ou obligatoire.⁵⁷ Au sein de chaque hôpital, ces systèmes ont été jugés utiles pour identifier, analyser et réduire le risque de survenue d’évènements de manière centralisée. L’objectif poursuivi n’est pas de condamner les professionnels responsables d’éventuelles erreurs de prise en charge, mais plutôt d’investiguer ces évènements pour en déterminer les causes racines et les éradiquer, en relayant largement les enseignements appris de ces analyses de terrain. S’apparentant à un véritable travail de ‘détective’, l’analyse de ces incidents cliniques doit être menée de manière systématique, en mettant l’accent sur les facteurs organisationnels plutôt que les individus.⁵⁸ En effet, bien

qu'une action isolée puisse être la cause immédiate d'un accident, une analyse plus approfondie révèle habituellement une série d'écarts à des pratiques sûres dont la survenue est influencée par l'environnement de travail (figure 3).

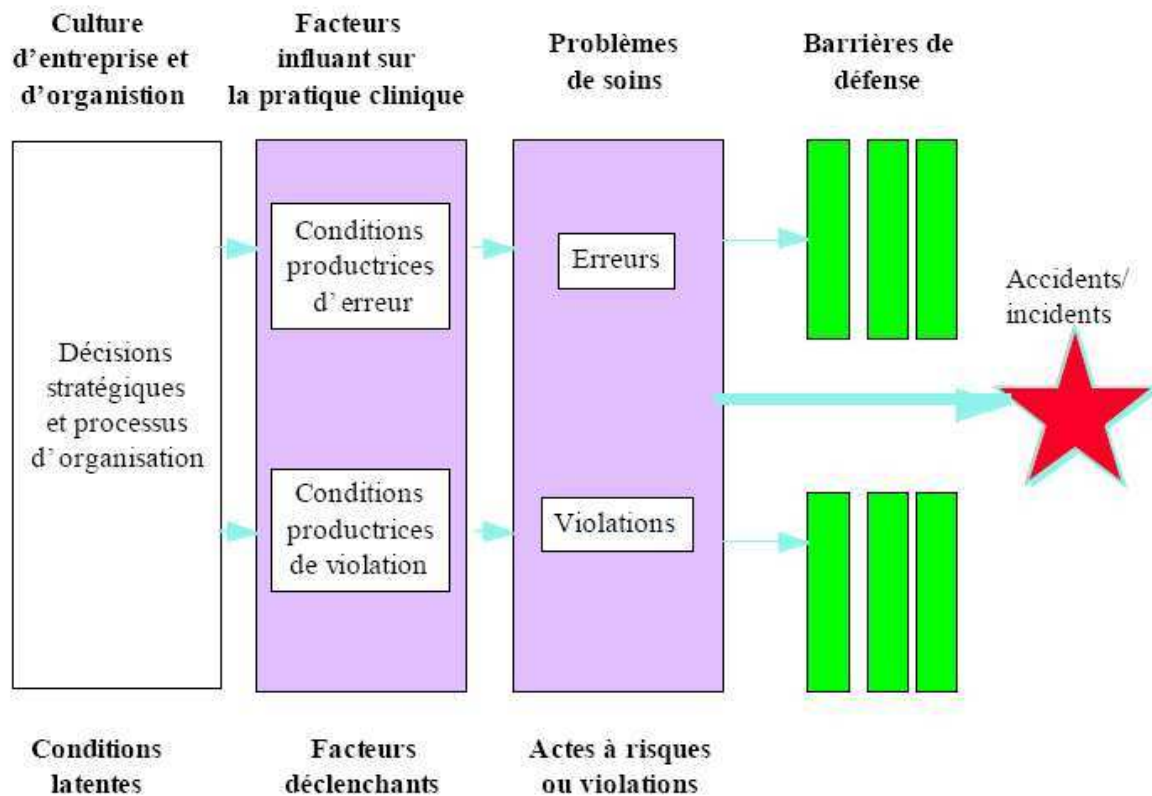


Figure 3. Modèle des causes organisationnelles d'accident⁵¹

La recherche et l'identification des causes de survenue d'un événement indésirable requièrent une enquête détaillée sur l'ensemble des facteurs pouvant concourir au résultat de la prise en charge. Sa mise en œuvre nécessite des moyens conséquents et un temps d'analyse suffisant pour être efficace. Une telle démarche peut s'avérer particulièrement ardue mais aussi fructueuse au sein d'un environnement aussi complexe que celui du bloc opératoire.

2.1.5. Garantir la sécurité du patient au bloc opératoire

La chirurgie constitue une part majeure des soins hospitaliers. On estime que 234 millions d'opérations chirurgicales importantes sont effectuées chaque année dans le monde, ce qui correspond à une intervention pour 25 habitants.⁵⁹ En France, plus de trois millions d'hospitalisations sont enregistrées annuellement pour réalisation d'un acte chirurgical.⁶⁰ Ces estimations n'incluent que les procédures menées en salle d'opération, impliquant incision, excision, manipulation ou suture de tissus et nécessitant généralement une anesthésie locale ou générale ou une sédation profonde.

Compte tenu de leur nombre important, la réduction des événements indésirables liés à ces interventions constitue pour tous les pays une priorité. Dans les pays industrialisés 3 à 16% des interventions chirurgicales entraînent des complications sérieuses, avec un taux d'incapacité permanente ou de mortalité allant de 0,4 à 0,8%.⁶¹ Dans les états du Colorado et de l'Utah, 2% des patients hospitalisés présenteraient un événement indésirable grave lié à leur passage dans un bloc opératoire, évitable dans 74% des cas.⁶² Un audit mené sur un échantillon de 144 établissements français a estimé la proportion d'événements indésirables graves survenus au cours d'un séjour hospitalier avec chirurgie à 2,9%. Le risque d'événement était plus élevé en chirurgie cardio-thoracique, en gynécologie et en chirurgie urologique qu'en ophtalmologie ou en chirurgie cervico-faciale.¹³ La fréquence de survenue d'une complication ou d'un décès après un acte chirurgical s'avère hautement variable selon la spécialité mais aussi d'un hôpital à l'autre.⁶³

Dans le cadre du lancement du programme '*Safe Surgery Saves Lives*' en 2008,⁶¹ l'OMS a fixé une série de dix objectifs à atteindre en vue d'améliorer la sécurité chirurgicale à travers le monde (table V). Une liste de six indicateurs visant à surveiller les soins chirurgicaux dispensés par chaque pays a également été proposée, incluant : le nombre de blocs opératoires disponibles, le nombre d'opérations réalisées, le nombre de chirurgiens et d'anesthésistes accrédités, la proportion de décès par jour de chirurgie et la proportion de décès postopératoires.⁶⁴ Enfin, une étude internationale 'avant-après' a été menée afin d'évaluer l'efficacité d'une *check-list* chirurgicale comprenant 18 items à vérifier avant l'induction anesthésique (*Time out* : vérification de l'identité, du site opératoire, des risques allergiques et hémorragiques), avant l'incision (*Sign in* : présentation des membres de l'équipe et revue systématique des risques) et avant la sortie du bloc opératoire (*Sign out* : résumé de l'intervention et des prescriptions postopératoires). Ses résultats ont montré une

réduction relative de 36% de la survenue d'évènements indésirables graves,⁶⁵ entraînant l'acceptation rapide de cette *check-list* dans de nombreux pays. En France, la HAS a depuis intégré l'utilisation d'une *check-list* chirurgicale à la procédure de certification des établissements de santé.⁶⁶

Table V. Dix objectifs essentiels pour une chirurgie plus sûre⁶¹

Objectif 1	L'équipe opérera le bon patient au bon endroit
Objectif 2	L'équipe aura recours à des méthodes connues pour prévenir les effets néfastes de l'administration d'anesthésiques, tout en protégeant le patient de la douleur
Objectif 3	L'équipe sera prête à faire face à une perte de fonction respiratoire susceptible de mettre en danger la vie du patient
Objectif 4	L'équipe sera prête à faire face au risque d'hémorragie massive
Objectif 5	L'équipe se gardera de provoquer une allergie ou des réactions indésirables à des médicaments connus pour présenter un risque important pour le patient
Objectif 6	L'équipe utilisera de manière conséquente des méthodes connues pour minimiser le risque d'infection chirurgicale
Objectif 7	L'équipe fera en sorte qu'aucune compresse ou instrument ne soit oublié par erreur dans la plaie après l'intervention
Objectif 8	L'équipe mettra en sécurité et identifiera avec précision tous les prélèvements chirurgicaux
Objectif 9	L'équipe communiquera de manière efficace et échangera les informations importantes pour effectuer l'intervention en toute sécurité
Objectif 10	Les hôpitaux et systèmes de santé publique établiront une surveillance en routine de la capacité, du volume et des résultats chirurgicaux

A l'hôpital, le bloc opératoire représente l'un des environnements de travail les plus exposé au risque de survenue d'évènements indésirables liés aux soins, en raison de la complexité des actes qui y sont réalisés et de la diversité des métiers impliqués (figure 4).⁶⁷ L'élément clé du soin chirurgical demeure certes l'opération en elle-même, mais l'établissement d'un diagnostic correct menant à une indication opératoire appropriée, la préparation du patient et sa surveillance ou la prescription adéquate de thérapeutiques participent également à son succès. Les erreurs potentiellement commises concernent donc l'ensemble de la période péri-opératoire et peuvent être attribuées à des facteurs structurels ou liés aux procédures de soins.⁴⁵

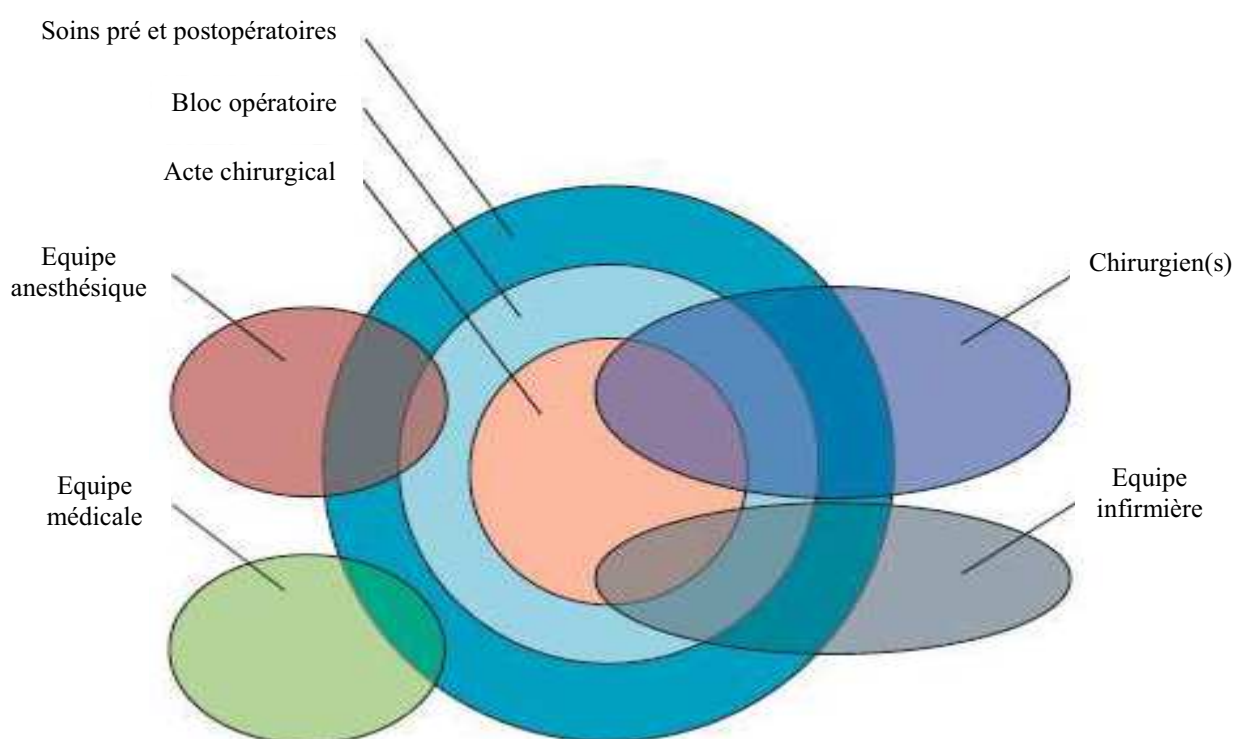


Figure 4. Représentation schématique de la complexité d'une intervention chirurgicale⁶⁷

Parmi les complications classiquement rencontrées lors d'une prise en charge chirurgicale, on retiendra particulièrement le décès, l'épisode hémorragique, l'infection du site opératoire,⁶⁸ les manifestations thromboemboliques,⁶⁹ l'erreur de côté,⁷⁰ l'oubli de compresses ou d'instruments.⁷¹ L'anesthésie peut également entraîner l'hypothermie, l'arrêt cardiorespiratoire ou l'ulcère de décubitus. Toutefois, l'hétérogénéité des modalités d'identification (méthode et délai de détection) et de collecte (exhaustivité et qualité du codage) de ces complications ainsi que leur ajustement difficile (caractéristiques du patient,

du chirurgien ou de l'hôpital), constituent souvent un obstacle à l'interprétation précise de leur variation dans le temps ou entre hôpitaux.⁷²

L'analyse des résultats chirurgicaux se limite fréquemment aux seules caractéristiques cliniques du patient (*case-mix*)⁷³ et à la compétence technique du chirurgien.⁷⁴ Par conséquent, une fois le résultat correctement ajusté sur le *case-mix*, sa variabilité est supposée s'expliquer par la performance individuelle de l'opérateur. Les caractéristiques liées au chirurgien, comprenant sa formation initiale, son expérience⁷⁵ ou son volume d'activité,^{76,77} expliquent effectivement une part de cette variabilité. Lors de l'introduction d'une technique chirurgicale innovante, la courbe figurant son apprentissage peut s'avérer très différente d'un chirurgien à l'autre⁷⁸ et requiert parfois une longue période avant d'être maîtrisée.⁷⁹ Cependant, loin de dépendre uniquement de l'habileté technique du chirurgien ou des caractéristiques du patient opéré, le résultat d'une chirurgie varie aussi selon d'autres facteurs, en partie liés aux conditions de travail. Charles Vincent a développé une approche conceptuelle visant à expliquer, de manière plus complète, les facteurs concourant à la sécurité du patient opéré incluant : la multiplicité des interfaces, la coordination d'équipe et le leadership, la culture organisationnelle et la qualité de la prise de décision. Il apparaît que le succès ou l'échec d'une chirurgie peut résulter d'une multitude de variables parfois difficiles à mesurer parmi lesquelles la sophistication des équipements utilisés, la quantité importante d'informations à gérer, la nature de la communication (écrite ou verbale) au sein de l'équipe ou le degré d'urgence et d'incertitude conditionnant la réalisation des soins.⁸⁰ Cette complexité, combinée avec la surcharge de travail, la fatigue et le stress des professionnels,⁸¹ rendent le soin chirurgical particulièrement propice à la survenue d'erreurs humaines. Une planification inadaptée des tâches¹³ et l'absence de supervision des plus jeunes⁸² y contribuent également, de même qu'une mauvaise communication^{83,84,85} ou une culture de la sécurité défaillante au sein des équipes.⁸⁶ En définitive, des causes liées au facteur humain et à une mauvaise organisation au sein du bloc seraient à l'origine de nombreux accidents évitables.⁸⁷ Le développement récent de l'analyse systémique des événements indésirables à l'hôpital résulte de ces constatations,⁸⁸ de même que la conception de *check-lists* visant à faciliter le travail en équipe multidisciplinaire.

2.2. La carte de contrôle

2.2.1. Apport du Dr Shewhart au contrôle qualité dans l'industrie

Walter Andrew Shewhart est né le 18 mars 1891 aux États-Unis. Après l'obtention d'un doctorat de physique à l'Université de Californie, il est embauché à la *Western Electric Company* en 1918. En 1925 il est transféré aux *Bell Telephone Laboratories* dont il sera salarié jusqu'à sa retraite en 1956, avant son décès le 11 mars 1967.

Le nom de Shewhart est associé aux méthodes de contrôle statistique de la qualité qui s'est considérablement développé au vingtième siècle dans l'industrie de grandes séries des produits manufacturés. Ses travaux sur les cartes de contrôle ont pour origine la volonté de la Western Electric d'investir dans une démarche d'assurance qualité dès 1922. Lorsque les *Bell Telephone Laboratories* sont créés par l'entreprise en 1925, l'ingénierie de l'inspection y a sa place dans l'organigramme et Shewhart en deviendra l'un des membres de la direction technique. Ainsi, il disposait d'une légitimité institutionnelle et de moyens importants pour intervenir sur de nombreux sites de production de l'entreprise et y expérimenter ses méthodes de contrôle avec des données réelles.⁸⁹ Accompagné d'un groupe d'ingénieurs et de statisticiens, il fut le premier à établir une théorie scientifique de la qualité, basée sur la relation étroite existant entre l'amélioration de la qualité et le contrôle de sa variation. Le postulat de départ de cette théorie était que si la différence se trouvait dans la nature, il en était de même pour les résultats des processus créés par l'Homme : la variation existait quoi que l'on puisse faire. Il en résultait que la qualité d'un processus de production était inversement proportionnelle à sa variabilité. Cette variabilité pouvait être suivie dans le temps à l'aide d'indicateurs censés refléter la qualité du produit final.

Shewhart a passé la majeure partie de sa carrière à étudier la variabilité de ces indicateurs, considérant qu'elle pouvait être riche d'enseignements sur les causes l'ayant provoquée. La détection puis l'identification de ces causes de variabilité devait permettre de mieux les contrôler en vue d'améliorer la qualité du produit final. Ainsi, l'application des méthodes statistiques garantissait un usage optimal des matières premières et des processus de production, permettant la réalisation d'économies de fabrication et de produits manufacturés de très haute qualité. Utilisée dans un premier temps pour construire des téléphones plus fiables, cette théorie scientifique de la qualité fut ensuite étendue à l'ensemble du monde industriel, une fois son efficacité reconnue. Depuis le siècle dernier, elle a été largement

acceptée et validée internationalement, particulièrement grâce à Edwards Deming qui l'a exportée pour permettre au Japon de combler son retard après la seconde guerre mondiale. Plus couramment appelée maîtrise statistique des processus, elle a été validée dans l'industrie et fait l'objet aujourd'hui de normes nationales (Afnor) et internationales (ISO) qui établissent ses conditions d'application et de mise œuvre.⁹⁰

Bien qu'il ait inventé la première carte de contrôle en 1924, c'est en 1931 que Shewhart publie l'ensemble de ses travaux dans un magistral traité qui demeure d'actualité.⁹¹ Shewhart avait constaté que les opérateurs réagissaient de manière excessive et procédaient à des changements inappropriés sur la chaîne de production en réponse à des variations d'indicateurs souvent aléatoires. De tels changements s'avéraient peu efficaces dès lors qu'ils introduisaient encore davantage de variabilité dans le processus de production, réduisant d'autant plus sa fiabilité.⁹² Guidée par la carte de contrôle, l'interprétation adéquate d'un indicateur devait apprendre à l'opérateur quand et comment réduire cette variabilité. La carte de contrôle s'est avérée particulièrement bien adaptée à un usage routinier dans les ateliers grâce à une exploitation ingénieuse et élégante des représentations graphiques, permettant de limiter le recours aux mathématiques pour leur interprétation. La qualité des produits étant définie par un ensemble de variables mesurables, supposées suivre des distributions statistiques, la carte de contrôle fournissait une représentation visuelle de l'évolution des paramètres de leur distribution (moyenne et dispersion). Les données déjà disponibles fournissaient un témoignage sur la base duquel l'établissement d'une prévision était possible assortie d'un degré de confiance.⁹³ Ainsi, la carte de contrôle permettait de repérer la survenue de causes significatives de perturbation (appelées 'assignables'), à l'origine d'une variation anormalement élevée de l'indicateur autour de la moyenne. L'investigation et l'élimination progressive de toutes les causes assignables devaient alors aboutir à l'obtention d'un processus de production sous contrôle statistique.

Shewhart a apporté une autre contribution majeure au contrôle qualité industriel par l'intégration de la carte de contrôle dans un modèle de management dédié à une amélioration continue de la qualité. Selon lui, processus de contrôle statistique et processus de production de connaissance étaient intimement liés selon trois phases successives : concevoir une hypothèse (spécification), mener l'expérience (production), puis tester l'hypothèse (inspection). Il a donc développé un cycle en quatre étapes se répétant indéfiniment :⁹⁴

- *Plan* : planifier ce qui peut être amélioré et les changements nécessaires ;
- *Do* : mettre en œuvre ces changements ;
- *Study* : interpréter le résultat des indicateurs et étudier le processus de production ;
- *Act* : en déduire des enseignements si les résultats escomptés ne sont pas atteints.

Ce cycle est davantage connu aujourd'hui sous l'appellation de 'roue de Deming' (figure 5), lequel a considérablement participé à la diffusion et à la vulgarisation de ce concept fondamental de la qualité moderne. La roue de Deming repose également sur 4 étapes : *Plan* (planifier, définir les objectifs et les moyens) - *Do* (agir, réaliser les tâches définies) - *Check* (vérifier, comprendre les résultats) - *Act* (améliorer, réagir).⁹⁵

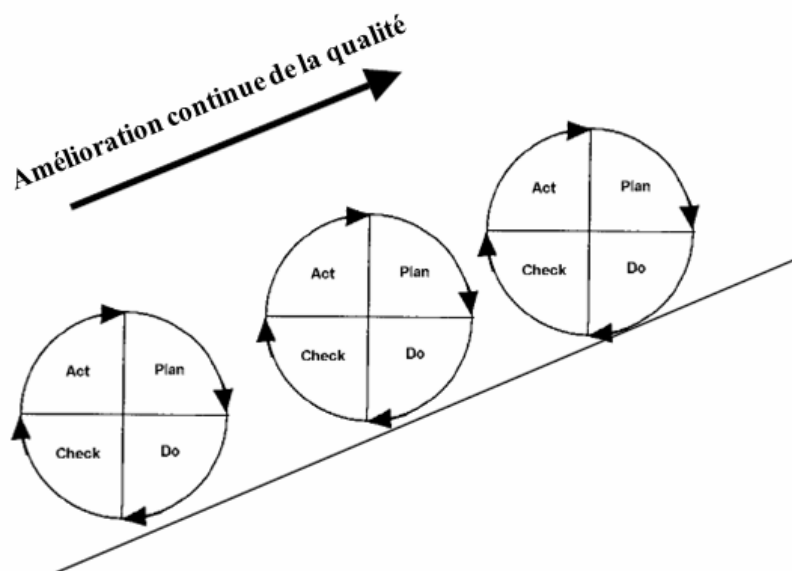


Figure 5. Représentation schématique de la roue de Deming

2.2.2. Principes d'application d'une carte de contrôle

La carte de contrôle est un outil graphique d'aide à la décision. Elle est spécifique d'un indicateur et autorise une analyse visuelle de sa mesure répétée dans le temps à partir d'échantillons prélevés périodiquement. Partant du principe qu'aucun processus ne peut aboutir à un résultat constant, la carte de contrôle vise à faciliter l'interprétation de ses variations. Pour y parvenir, les causes de variabilité du résultat final d'un processus sont regroupées en deux types :

- Les causes communes de variation qui sont constantes, structurelles, inhérentes au processus étudié. On considère que ces causes sont dues au hasard et qu'elles ne sont ni identifiables, ni éliminables.
- Les causes spéciales (ou assignables) de variation qui sont irrégulières, inhabituelles, non aléatoires et pour lesquelles une étiologie peut être retrouvée. Il s'agit de sources de variabilité que l'on peut identifier et éliminer, une fois que la relation entre leur survenue et le résultat obtenu a été mise en évidence.⁹⁶

En permettant de faire la distinction entre ces deux types de causes, la carte de contrôle permet de catégoriser la variation d'un indicateur en fonction de l'action qui doit être mise en œuvre pour la réduire. En effet, à l'inverse des causes communes, les causes spéciales requièrent des investigations approfondies car elles témoignent d'une variation importante du processus.⁹⁷ Pour parvenir à les détecter et à déterminer si le processus est sous contrôle, uniquement soumis à l'influence de causes communes, trois courbes sont figurées horizontalement sur la carte. La première est appelée la ligne centrale et représente le niveau de référence de l'indicateur, habituellement sa valeur moyenne calculée sur l'ensemble de la période de suivi. Les deux autres sont nommées les limites de contrôle inférieure et supérieure et représentent les seuils de variabilité acceptable de l'indicateur. Ces limites sont classiquement tracées à une distance de 3 déviations standard (DS) par rapport à la valeur de la ligne centrale.⁹⁸ Lorsqu'un des points figurant la variation de l'indicateur dans le temps est positionné en dehors de ces limites, cela signifie qu'une cause spéciale de variation a été détectée et par conséquent que le processus n'est plus sous contrôle (figure 6). Cette règle de décision garantit le meilleur équilibre entre la sensibilité de l'outil pour détecter une cause spéciale (c'est-à-dire à éviter de considérer une cause spéciale comme étant une cause commune) et sa spécificité pour éviter de déclencher de fausses alarmes (c'est à dire à éviter de considérer une cause commune comme étant une cause spéciale).⁹⁹ Deux autres limites

d'alerte, inférieure et supérieure, peuvent également être représentées, à une distance de 2 déviations standard par rapport à la valeur de la ligne centrale afin d'améliorer la sensibilité de détection de la carte. Dans ce cas, la survenue d'une cause spéciale est également signalée par la présence de deux points sur trois consécutifs entre une limite d'alerte (2DS, 95,5%) et une limite de contrôle (3DS, 99,7%), d'un même côté de la ligne centrale. L'ajout de ces limites d'alerte expose toutefois à une augmentation du risque de fausses alarmes.⁹⁷

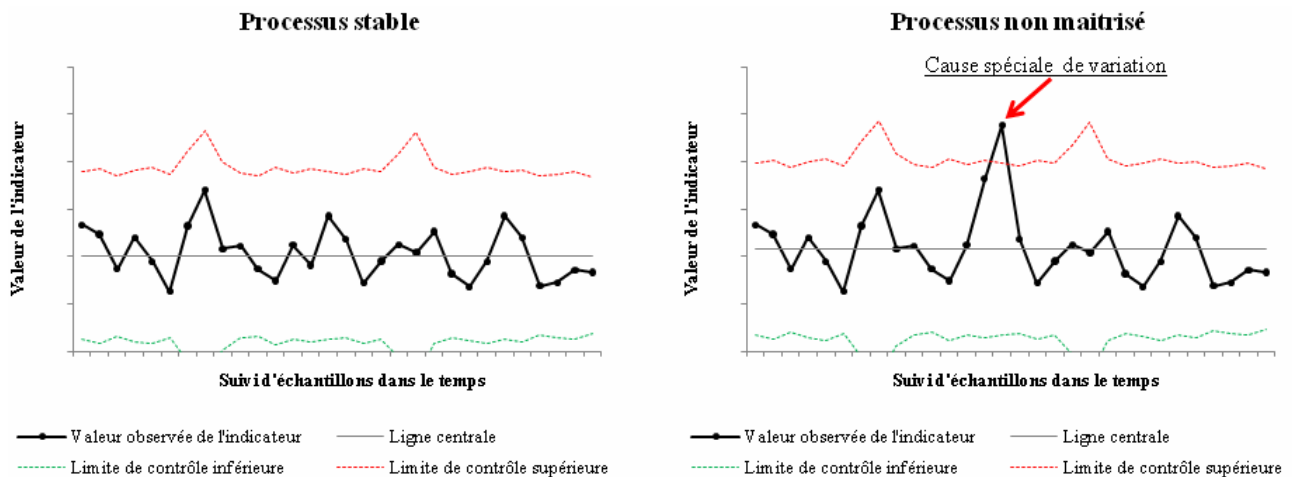


Figure 6. Exemple de carte de contrôle sur processus stable ou non maîtrisé

L'objectif poursuivi par l'utilisation d'une carte de contrôle est double pour amener le processus de production au niveau de qualité requis et l'y maintenir. En premier lieu, il faut réduire la variabilité des résultats observés sur l'indicateur afin de parvenir progressivement à un processus stable. Cela implique la détection et l'identification systématique de toutes les causes spéciales pouvant survenir et rendre le processus imprévisible. Selon la limite franchie (inférieure ou supérieure), il s'agira soit d'une cause spéciale dégradant la qualité du processus et donc à éliminer, soit d'une cause spéciale en améliorant la qualité et donc à conserver. Il faut néanmoins préciser que, si la carte de contrôle permet la détection d'une cause spéciale, elle n'en identifie pas l'origine. Pour mener cette enquête, un journal de bord est classiquement tenu à jour dans lequel est consigné l'ensemble des modifications ayant pu affecter le processus et en influencer les résultats. D'autres outils du contrôle qualité peuvent également être utilisés pour identifier l'origine des causes spéciales, tels les diagrammes de Pareto ou d'Ishikawa, le *flow-chart*, etc.⁹⁰

Dans un second temps, une fois le processus maîtrisé, on peut envisager d'en diminuer la variabilité résiduelle due aux causes communes, si celle-ci s'avère trop importante ou que

le niveau de qualité atteint est jugé insatisfaisant. Dans ce cas, une réorganisation complète du processus doit être menée en visant la simplicité, un système simple étant plus fiable qu'un système complexe qui multiplie les intermédiaires et donc la probabilité d'erreurs, sources de variation incontrôlée.⁹² Des actions d'amélioration peuvent également être expérimentées à l'aide de la carte de contrôle, seul un système stable (duquel on a vérifié l'absence de cause spéciale) et prédictible garantissant leur évaluation rigoureuse.¹⁰⁰

2.2.3. Différents types de cartes applicables en chirurgie

Initialement confinée au contrôle qualité dans l'industrie, l'usage de cartes de contrôle s'étend désormais au champ de la santé. Dès 1942, Edwards Deming recommandait l'implémentation de tels outils pour suivre et réduire la survenue d'événements indésirables liés aux soins.¹⁰¹ On peut situer l'origine de l'utilisation de la maîtrise statistique de processus en santé à l'occasion de l'inauguration du *National Demonstration Project in Quality Improvement in Health Care* à Boston (États-Unis) en 1987.⁹⁰ Les participants (des médecins, des directeurs d'hôpitaux et des responsables qualité du secteur industriel) y étaient réunis pour discuter l'intérêt d'une application des méthodes qualité de l'industrie (Ford, NASA, IBM, etc.) aux soins. Ce travail concluait à l'intérêt de ces techniques pour l'amélioration de la qualité en santé.¹⁰² Aujourd'hui la *Joint Commission on Accreditation of Healthcare Organizations* associe l'utilisation de la maîtrise statistique de processus au cycle *Plan-Do-Check-Act* dans le cadre de la démarche qualité des hôpitaux américains.¹⁰³

Cartes de Shewhart

Les cartes les plus simples à utiliser et les plus connues demeurent les cartes de contrôle décrites en 1924 par Shewhart et qui portent son nom. Il en existe deux grands types selon la nature de l'indicateur que l'on souhaite suivre et sa loi de distribution statistique. Les cartes de contrôle 'aux mesures' permettent de suivre une variable quantitative reposant sur une loi de distribution normale (table VI). Dans ce cas, chaque carte comportera deux graphiques : un pour le suivi de la tendance centrale de l'indicateur et un autre pour le suivi de sa dispersion. Les cartes aux mesures les plus utilisées sont : la carte (\bar{X}, s) à moyenne - écart-type et la carte (\bar{X}, R) à moyenne - étendue. Les cartes de contrôle 'aux attributs' permettent de suivre une variable qualitative. Dans ce cas, la carte ne comporte qu'un seul graphique représentant le nombre ou le pourcentage de non-conformes (selon que l'échantillon est de taille fixe ou non) ou de non-conformités (selon qu'une même unité statistique de l'échantillon peut présenter plusieurs événements). Elles font appel à une loi de distribution de probabilité binomiale ou de Poisson. Les principales cartes de contrôle aux attributs sont : la carte p pour le suivi de la proportion de non-conformes, la carte np pour le suivi du nombre de non-conformes, la carte c pour le suivi du nombre de non-conformités et la carte u pour le suivi du taux de non-conformités. Depuis Shewhart, le contrôle qualité a été enrichi par d'autres types de cartes de contrôle. La carte d permet de suivre des non-conformités mais en attribuant un poids différent selon la gravité de chaque type de non-conformité. L'indicateur

suivi est alors un indice de qualité qui est la moyenne pondérée des non-conformités. Les cartes *g* et *h* utilisent une loi de distribution de probabilité géométrique pour des indicateurs dont la variable a une fréquence d'apparition très faible. Plutôt que d'utiliser le nombre de cas on utilise le délai entre l'occurrence de deux cas successifs.¹⁰⁴

Table VI. Différents types de cartes de Shewhart

Type de carte	Distribution	Indicateur	Exemples
<i>X - S</i>	Normale (Gaussienne)	Moyenne par échantillon avec écart-type correspondant	Durée de séjour Données physiologiques
<i>np</i>	Binomiale	Nombre total d'évènements par échantillon de taille constante (un seul événement est possible pour chaque unité de l'échantillon)	Nombre d'infections du site opératoire Nombre de réadmissions
<i>p</i>	Binomiale	Proportion d'évènements par échantillon de taille variable (un seul événement est possible pour chaque unité de l'échantillon)	Proportion d'infections du site opératoire Proportion de réadmissions
<i>c</i>	Poisson	Nombre total d'évènements par échantillon de taille constante (un événement ou plus est possible pour chaque unité de l'échantillon)	Nombre de chutes Nombre d'infections sur cathéter
<i>u</i>	Poisson	Taux d'évènements par échantillon de taille variable (un événement ou plus est possible pour chaque unité de l'échantillon)	Nombre moyen de chutes pour 100 jours-patient Nombre d'infections sur cathéter pour 100 jours-cathéter
<i>g</i>	Géométrique	Nombre de cas ou délai de survenue entre les non-conformes	Nombre d'interventions entre les infections Nombre de patients entre les complications

Ces différents types de cartes de contrôle ont été expérimentés dans de nombreuses spécialités médicales et structures de soins, se montrant applicables dans divers contextes de prises en charge.¹⁰⁵ Dans le cadre du contrôle qualité, l'utilisation en routine de la carte de contrôle est utile pour interpréter et réduire les sources de variabilité liées aux soins. En convertissant les données recueillies en connaissance, elle guide les équipes multidisciplinaires vers l'action la plus appropriée dans une situation donnée afin d'améliorer en continu la qualité des prises en charge.¹⁰⁶ Dans le champ de la recherche sur les services de santé, elle représente une méthode robuste et peu coûteuse pour évaluer l'efficacité des

changements réalisés sur les soins.¹⁰⁷ Reposant sur un schéma pseudo-expérimental, elle constitue une approche complémentaire aux essais randomisés pour montrer l'impact d'une action visant à améliorer la sécurité du patient.¹⁰⁸ Les cartes de contrôle ont été également utilisées avec succès dans d'autres domaines, tels les systèmes de surveillance hospitaliers¹⁰⁹ ou communautaires¹¹⁰ ou le suivi individuel de l'état de santé du patient.¹¹¹

Toutefois, à l'origine ces cartes de contrôle ont été spécifiquement développées pour l'analyse de processus de production industriels. Quand on souhaite les appliquer au suivi des complications chirurgicales, deux contraintes majeures peuvent en limiter l'intérêt. La première concerne le risque préopératoire des patients. Chaque patient a des caractéristiques propres pouvant influencer le résultat de sa prise en charge quelle que soit la qualité de réalisation du geste par le chirurgien. Par conséquent, la variabilité du risque préopératoire entre patients requiert un ajustement adéquat afin de pouvoir suivre la performance des chirurgiens indépendamment de ce facteur. Peu de travaux ont été conduits dans ce domaine, amenant à la construction de cartes ajustées sur le *case-mix* des patients opérés et dont les conclusions se sont avérées contradictoires.^{112,113,114} La seconde contrainte concerne la rareté des complications opératoires, laquelle peut réclamer une sensibilité de détection accrue de la carte pour analyser des variations de faible amplitude sur l'indicateur. Cela est d'autant plus vrai lorsque l'on souhaite suivre la performance individuelle d'un chirurgien. Les cartes de contrôle classiques de Shewhart sont basées sur une approche populationnelle permettant de signaler de larges variations sur des échantillons de cas groupés. Mais elles ne prennent en compte que l'information obtenue lors du dernier échantillon prélevé pour représenter l'indicateur. Afin de détecter de petits changements lors du suivi des complications opératoires, d'autres types de cartes ont été développés. Appelées cartes de séries cumulées, elles fournissent un suivi en temps réel du résultat cumulé (succès ou échec) de chaque procédure.¹¹⁵

Cartes de séries cumulées (*CUSUM charts*)

Parallèlement à la forte croissance de la littérature médicale sur les cartes de séries cumulées,¹¹⁶ beaucoup de confusion persiste sur leur dénomination exacte et leurs modalités d'utilisation (table VII et figure 7).^{117,118} Dans l'ensemble, la construction et l'interprétation de ce type de cartes s'avèrent plus complexes que pour les cartes de contrôle traditionnelles de Shewhart. Certaines cartes permettent une interprétation intuitive de la performance des

chirurgiens tandis que d'autres proposent des limites autorisant le test formel d'une hypothèse. Certaines cartes sont destinées à la surveillance en routine de la performance des chirurgiens,¹¹⁹ d'autres étant plus adaptées pour établir ponctuellement leurs courbes d'apprentissage (*learning curve*).¹²⁰ Enfin, certaines cartes reposent sur le suivi de données brutes, d'autres sur des modèles d'ajustement relatifs aux caractéristiques des patients.

Table VII. Différents types de cartes de séries cumulées

Type de carte	Utilisation		Interprétation	
	Apprentissage	Monitoring	Limite(s)	Test statistique
<i>Cumulative Failure graph</i>	Visualisation de la courbe	Visualisation de la courbe	Pas de limites	Simple mais peu informatif, pas de test
<i>Observed minus Expected</i>	Visualisation de la courbe	Visualisation de la courbe	Pas de limites	Très intuitif, pas de test
<i>SPRT</i>	Visualisation de la courbe	Position de la courbe par rapport aux limites	Limites inférieure (acceptable) et supérieure (inacceptable)	Peu intuitif mais test
<i>CUSUM test</i>	-	Position des deux courbes par rapport aux limites	Limites inférieure (amélioration) et supérieure (détérioration)	Peu intuitif mais test
<i>LC-CUSUM</i>	Position de la courbe par rapport à la limite	-	Limite inférieure de passage sous contrôle	Peu intuitif mais test

Le *cumulative failure graph* a l'avantage d'une carte très simple à construire mais qui n'apporte que peu d'information.¹²¹ L'*Observed minus Expected chart* (carte O-E)¹²² est une carte qui permet de visualiser tôt l'amorce d'une variation mais ne permet pas d'interprétation basée sur un test statistique. L'*Expected minus Observed chart*, qui représente une variante de la carte O-E, est dénommé *Variable Life Adjusted Display* (VLAD)¹²³ ou *Cumulative Risk Adjusted Mortality* (CRAM)¹²⁴ selon les équipes. Le *Sequential Probability Ratio Test* (SPRT) a été développé par Wald en 1945¹²⁵ et bénéficie de l'apport de limites de détection.¹²⁶ Néanmoins, l'approche empruntée pour le calcul de ces limites demeure controversée¹²⁷ et il convient de préférer l'utilisation de la *Cumulative Sum chart* (carte CUSUM),¹²⁸ introduite par Page en 1954, si l'on souhaite disposer d'un véritable test statistique permettant la détection d'une variation significative des résultats. La statistique suivie par la carte CUSUM est une somme d'écarts cumulés entre la moyenne de l'échantillon et une valeur cible. Essentiellement basée sur le dérèglement de la moyenne de l'indicateur, cette

carte permet d'augmenter artificiellement la taille de l'échantillon prélevé. Ainsi, elle détecte plus finement de petites tendances qui perdurent dans le temps¹²⁹ mais se révèle moins efficace pour détecter de grands dérèglages ponctuels. Deux types de cartes CUSUM coexistent aujourd'hui : le *CUSUM test*¹³⁰ utile au suivi de la performance en routine des chirurgiens et le *CUMulative SUMmation test for Learning Curves* (test LC-CUSUM)¹³¹ idéal pour tracer la courbe d'apprentissage d'une nouvelle technique chirurgicale. Dans les deux cas, la construction des limites s'avère complexe mais rigoureuse statistiquement pour y appliquer un test. Une carte (I,mR) aux valeurs individuelles - étendue mobile ou *Exponentially Weighted Moving Average* (EWMA), relativement proche de la carte CUSUM, a également été expérimentée. Cette carte utilise une statistique cumulant des moyennes pondérées d'échantillons issus de tous les prélèvements passés.¹³²

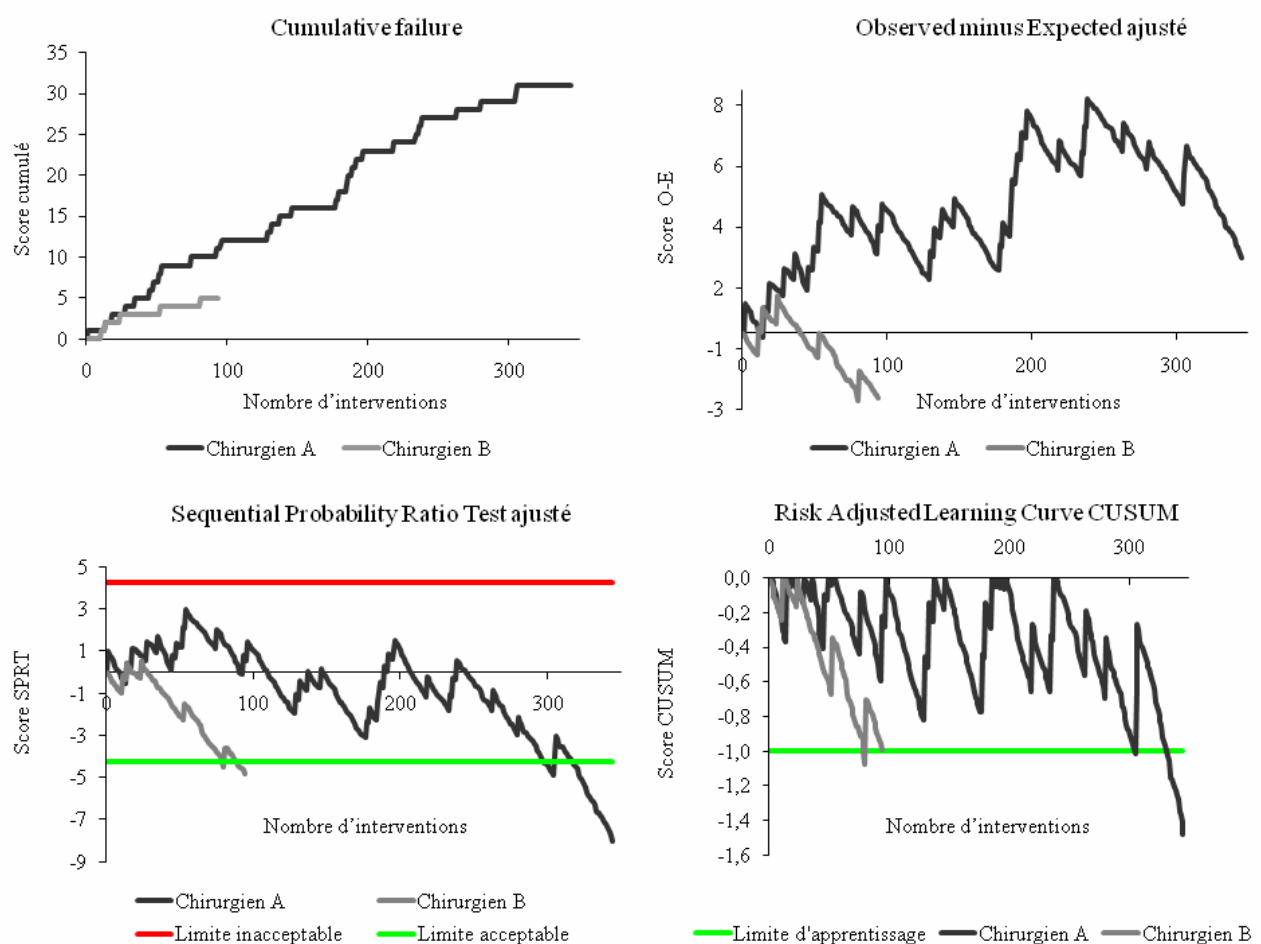


Figure 7. Cartes de séries cumulées représentant la courbe d'apprentissage individuelle d'une nouvelle technique par deux chirurgiens

En pratique, deux phases sont parfois distinguées pour l'application des cartes de contrôle dans l'industrie. Lors de la phase I, des données historiques sont utilisées pour fournir un état des lieux initial, estimer les indicateurs décrivant le comportement du processus de production et évaluer sa stabilité en détectant d'éventuelles causes spéciales de variation. La phase II consiste à suivre de manière prospective les indicateurs mesurés à partir d'échantillons de données prélevés périodiquement. L'application de cartes de contrôle de type Shewhart est fortement recommandée en phase I, tandis que leur utilisation combinée avec celle de cartes CUSUM s'avère pertinente en phase II.¹³³

La lecture des propriétés de chacune des cartes suggère une utilisation des cartes de contrôle de type Shewhart lorsque l'on suit les résultats d'une équipe de chirurgien, à la condition de s'intéresser à des complications opératoires suffisamment fréquentes. A l'inverse, pour le suivi individuel des résultats d'un même chirurgien ou la détermination de sa courbe d'apprentissage, le recours aux cartes de séries cumulées de type O-E ou CUSUM paraît plus approprié. Dans tous les cas, la carte développée sera longitudinale et visera à interpréter dans le temps les variations observées sur un indicateur de résultat qui devra être pertinent cliniquement, fiable et valide.

Carte transversale (*funnel plot*)

Un dernier type de carte a été développé pour mener des comparaisons de performance inter-hospitalières. Face aux tentatives récentes de la presse grand public pour classer les hôpitaux, certains systèmes de santé étrangers ont proposé une alternative. Les expériences conduites pour essayer de normaliser les classements à l'aide de *league tables* s'étant révélées décevantes,¹³⁴ des cartes de contrôle transversales,¹⁰⁶ ou *funnels plots*,¹³⁵ ont été conçues à la lueur des travaux de Shewhart. Dans ce cas, les auteurs ont cherché à identifier les hôpitaux s'écartant significativement d'une performance moyenne calculée pour l'ensemble des établissements et jugée comme acceptable. L'intérêt de la carte repose alors sur sa capacité à discriminer les hôpitaux hors-norme (*outliers*), c'est-à-dire ceux très performants ou peu performants (figure 8). Ainsi, il apparaît plus profitable de cibler les investigations sur ces *outliers*, afin d'identifier précisément les raisons de résultats excellents et les moyens utilisés pour y parvenir, ou au contraire les problèmes à l'origine de résultats inférieurs à la moyenne et les remèdes à apporter pour y remédier.¹¹⁵

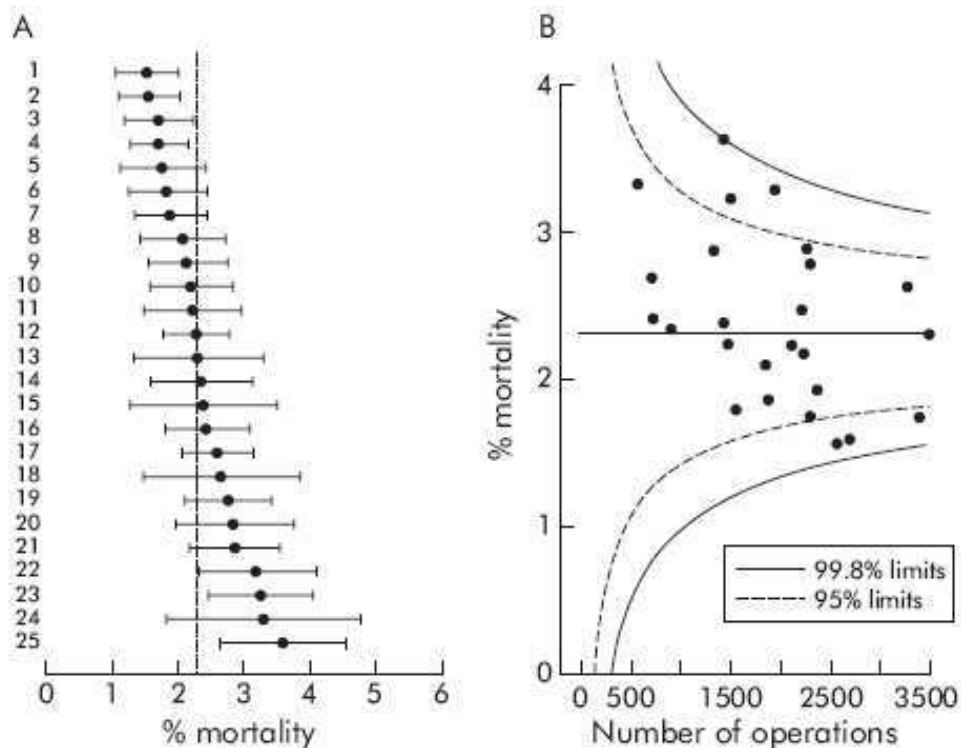


Figure 8. Mortalité à 30 jours après pontage coronarien dans 25 hôpitaux anglais
A. Forrest plot comparant l'intervalle de confiance à 95% au taux moyen global.
B. *Funnel plot* des taux observés selon le nombre d'interventions réalisées par établissement.
[Extrait de D.J. Spiegelhalter]¹³⁶

2.3. La chirurgie de la thyroïde

2.3.1. La pathologie thyroïdienne : rappel anatomo-physiologique et indications opératoires

La thyroïde est une petite glande située à la base du cou, sous le larynx et entourant les faces antérolatérales de l'extrémité supérieure de la trachée. Elle est souvent comparée à un papillon car constituée de deux lobes latéraux (droit et gauche) se rejoignant par un isthme médian. Un lobe pyramidal peut également exister à sa partie supérieure. Les deux lobes latéraux sont de forme ovale, d'approximativement 5 cm de long et 2 à 3 cm de large. Chacun est bordé par la trachée et l'œsophage sur sa face médiane, par les vaisseaux carotidiens sur ses faces postéro-latérales et par le muscle sternocléidomastoïdien sur sa surface antérolatérale. La thyroïde, qui est entièrement recouverte d'une fine capsule, est notamment vascularisée de chaque côté par des artères et veines thyroïdiennes supérieures et inférieures.

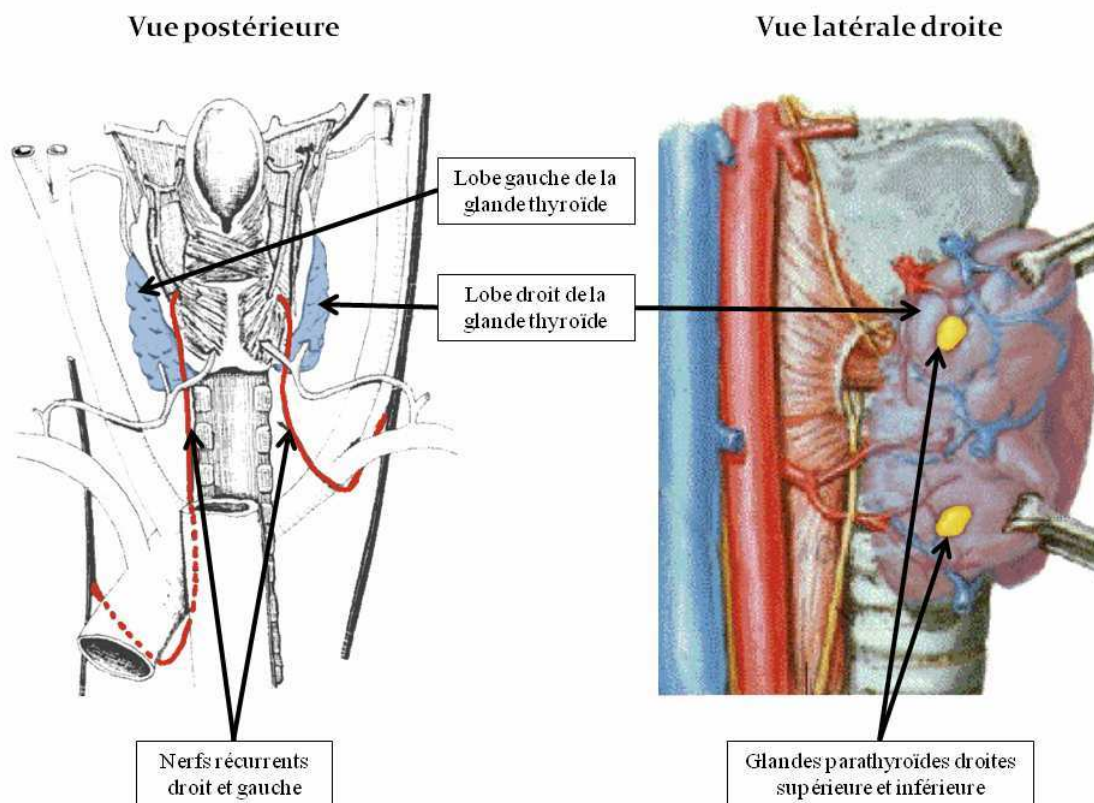


Figure 9. Représentation anatomique de la glande thyroïde

De part et d'autre de la glande thyroïde, deux branches du nerf vague sont essentielles pour permettre la mobilité adéquate des cordes vocales : la branche externe du nerf laryngé supérieur, se situant à proximité de l'artère thyroïdienne supérieure, et le nerf laryngé inférieur (ou nerf récurrent) qui passe derrière la thyroïde.¹³⁷ Quatre glandes parathyroïdes de 4 à 6 millimètres de grand axe, sont également étroitement accolées aux faces latérales de la glande de manière symétrique, même si leur nombre comme leur position peuvent varier grandement d'un individu à l'autre.¹³⁸ Leur sécrétion de parathormone permet de réguler la calcémie et chacune dispose généralement de sa propre vascularisation (figure 9).

Bien que la thyroïde ne pèse habituellement qu'une vingtaine de grammes, les hormones thyroïdiennes qu'elle secrète sont nécessaires à la croissance et au métabolisme de l'être humain. Il s'agit de la thyroxine (T4) et de la triiodothyronine (T3) qui, une fois libérées dans la circulation sanguine, se lient à des protéines porteuses dont la principale est la TBG (thyroxin binding globulin). Une part minoritaire de ces hormones circule sous forme libre. Ce sont les hormones T4 et T3 libres, biologiquement actives au niveau des tissus cibles. La synthèse des hormones thyroïdiennes est contrôlée par la thyroïdostimuline (TSH, *thyroid-stimulating hormone*), produite par les cellules thyroïdotropes de l'hypophyse, elle-même sous la dépendance de la thyrolibérine (TRH, *thyroid-releasing hormone*) synthétisée par l'hypothalamus. La synthèse et la libération de TRH et de TSH sont commandées par le niveau circulant des hormones thyroïdiennes (rétrocontrôle négatif) : toute chute de leur concentration dans l'organisme est immédiatement détectée par l'hypothalamus et l'hypophyse qui répondent en produisant la TRH et la TSH afin de stimuler la production hormonale de la thyroïde.¹³⁹

Les affections de la glande thyroïde sont fréquentes.¹⁴⁰ Deux cents millions de personnes, habituellement des femmes, présentent une anomalie thyroïdienne dans le monde. Ces anomalies peuvent être fonctionnelles (hyperthyroïdie ou hypothyroïdie), morphologiques (goitre, nodule) ou inflammatoires (thyroïdite).¹⁴¹ Les dysthyroïdies sont le plus souvent asymptomatiques ou pauci-symptomatiques.¹⁴² La prévalence de l'hypothyroïdie, parmi les populations bénéficiant d'un apport alimentaire suffisant en iode, est de 5‰ habitants pour les formes cliniquement évidentes et de 15‰ habitants pour les formes infracliniques.¹⁴³ Elle se manifeste classiquement par les signes suivants : lenteur des réflexes, frilosité, œdème papillaire, bradycardie, hyposudation, prise de poids, paresthésies, constipation, ralentissement moteur, voix rauque, hypoacousie, peau sèche, épaisse et

froide.¹⁴⁴ La prévalence de l'hyperthyroïdie est de 2‰ habitants pour les formes cliniquement évidentes et de 6‰ habitants pour les formes infracliniques.¹⁴³ Les manifestations liées à la thyrotoxicose regroupent typiquement : amaigrissement avec appétit conservé, asthénie musculaire, thermophobie, hypersudation, polydipsie, tachycardie, diarrhée, tremblement d'attitude, nervosité et agitation.¹⁴⁵ Le cancer de la thyroïde représente respectivement 0,5% et 1,5% de l'ensemble des cancers chez l'homme et la femme dont 94% de cancer différenciés avec un pronostic généralement excellent.¹⁴⁶ Son incidence tend à augmenter en France comme aux États-Unis depuis une trentaine d'années.^{147,148,149}

La majorité des pathologies thyroïdiennes ne nécessitent qu'une simple surveillance ou un traitement médical. Une intervention chirurgicale est indiquée dans certains cas comprenant le cancer de la thyroïde, la présence de symptômes obstructifs ou d'un goitre rétrosternal, une hyperthyroïdie récurrente ou ne répondant pas au traitement médical, des raisons esthétiques ou en raison de l'anxiété du patient.¹⁵⁰ La principale précaution à respecter est de n'intervenir chirurgicalement qu'après avoir posé une indication indiscutable et adaptée.¹⁵¹ Une chirurgie unilatérale, reposant sur la lobectomie ou la lobo-isthmectomie, sera privilégiée dans le cadre d'un nodule thyroïdien bénin unilatéral, sinon l'exérèse bilatérale de la glande sera réalisée.¹⁵² Une chirurgie bilatérale s'avère particulièrement indiquée en cas de processus pathologique affectant les deux lobes de la thyroïde ou bien quand le risque de récurrence est important comme dans le goitre multinodulaire bénin, la maladie de Basedow ou le cancer. La thyroïdectomie totale ou subtotale (qui est en fait une lobectomie totale ou subtotale bilatérale réalisée pendant la même opération) représente le traitement de choix pour les goitres toxiques, volumineux, anciens, inesthétiques ou évolutifs.¹⁵³ En cas de cancer différencié supra-centimétrique, une thyroïdectomie totale, souvent élargie avec curage ganglionnaire, doit être pratiquée.¹⁵⁴ Après une exérèse totale ou subtotale de la thyroïde, la prescription d'un traitement substitutif par hormones thyroïdiennes s'impose.¹⁴⁰

En 2008, 49 658 séjours pour prise en charge chirurgicale d'une affection thyroïdienne ont été réalisés dans les hôpitaux français. L'âge médian des patients opérés était de 52 ans et 79,2% étaient des femmes. La pathologie thyroïdienne sous-jacente était habituellement un goitre (54% des cas), un nodule (16%) ou un cancer (14%). L'acte réalisé sur la thyroïde était unilatéral dans 26% des cas (lobo-isthmectomie) alors qu'il s'agissait d'une thyroïdectomie bilatérale dans 74% des cas (dont 66% de thyroïdectomies totales). La majorité des patients étaient opérés dans un établissement public (59%). Au total, 805 établissements participaient à cette activité, 71% d'entre eux réalisant au moins 10 séjours par an, 15% au moins 100 séjours et 1% au moins 500 séjours (figure 10).

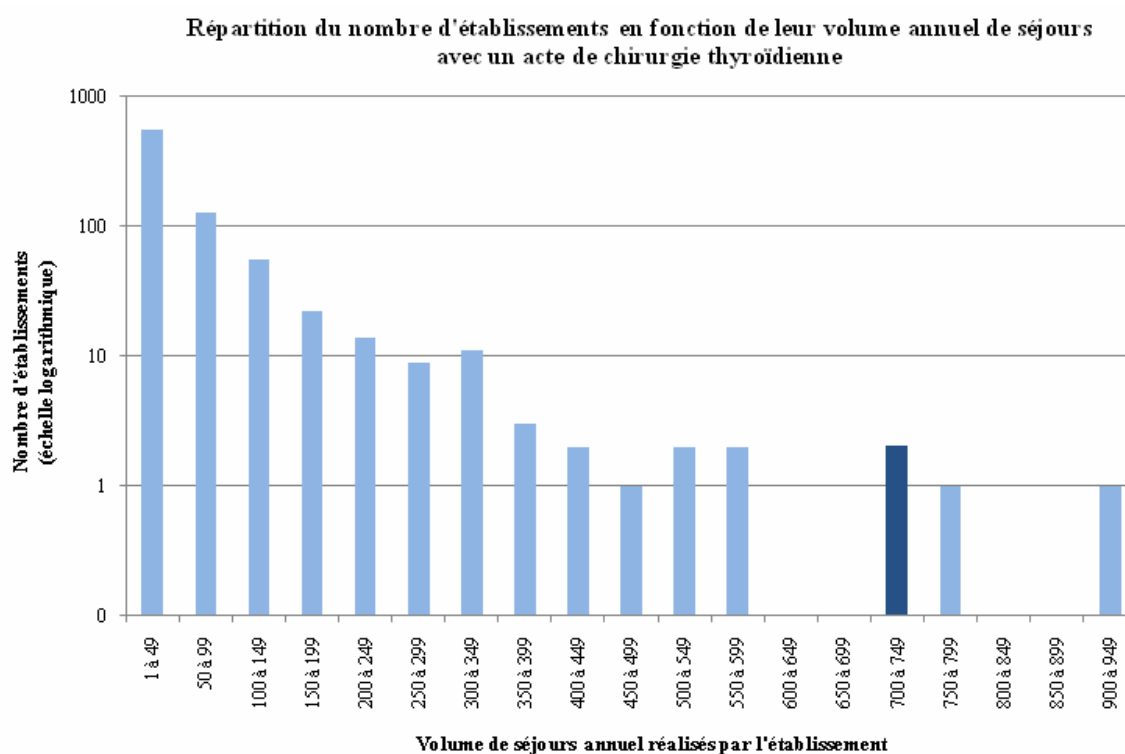


Figure 10. Répartition de l'activité de chirurgie thyroïdienne selon le volume d'activité des hôpitaux français (en foncé, catégorie d'appartenance du Centre Hospitalier Lyon Sud)

2.3.2. La thyroïdectomie : bref historique et conduite à tenir chirurgicale

Les goîtres de la glande thyroïde étaient déjà connus dans l'antiquité et décrits dans la littérature chinoise dès 2 700 avant JC. Au cours des douzième et treizième siècles, l'école de Salerne en Italie les réséquait à l'aide d'instruments inadaptés dans des conditions de sécurité douteuses. En 1821, Hedenus a rapporté l'exérèse de 6 goîtres suffocants par dissection et ligature de tous les vaisseaux, malgré une issue souvent fatale pour le patient. Avant la seconde moitié du dix-neuvième siècle, compte tenu du risque encouru par les patients (approximativement 20% de décès, 25% de paralysies des cordes vocales et 15% d'hypoparathyroïdies secondaires), les indications chirurgicales se limitaient aux situations mettant en péril la vie du patient et privilégiaient la réalisation d'opérations peu invasives.¹⁵⁵ Ainsi, en 1850 l'académie française de médecine condamnait la pratique d'opérations sur la thyroïde. Aux États-Unis, Greene considérait en 1871 que cette chirurgie devait être réservée aux cas les plus sévères et non pour le seul confort du patient.¹⁵⁶

La thyroïdectomie large ou totale fut introduite dans le traitement de la maladie de Basedow par Tillaux dès 1880. Néanmoins, les premières tentatives furent rapidement abandonnées en raison des dangers auxquels elles exposaient les patients, principalement le décès, les lésions récurrentielles, le myxoedème ou la tétanie. L'abstentionnisme des médecins vis-à-vis des Basedowiens fut donc la règle et la thyroïdectomie rarement réalisée jusqu'à la fin du dix-neuvième siècle, avant que Billroth à Vienne et surtout Kocher à Berne n'en révolutionnent l'approche, bénéficiant des avancées de l'anesthésie, l'antisepsie et l'hémostase.¹⁵⁷ Récompensé par le prix Nobel de médecine en 1909 pour ses travaux sur la physiopathologie et la chirurgie de la thyroïde, Kocher est souvent considéré comme le père de la chirurgie thyroïdienne.¹⁵⁸ Les améliorations successives qu'il a su apporter à sa pratique chirurgicale quotidienne ont permis une réduction notable de mortalité chez ses patients, passant de 14% en 1884 à 2,4% en 1889 et 0,18% en 1898.¹⁵⁹ A Lyon, Léon Bérard estimait en 1936 que « *la thyroïdectomie subtotale décidée opportunément et après une préparation logique des malades ne devait plus dépasser 4 à 5% de mortalité* ». ¹⁶⁰

Conjointement, la morbidité attribuable à la chirurgie thyroïdienne a diminué, la rendant plus sûre pour traiter à la fois les pathologies bénignes et malignes de la thyroïde. La dissection capsulaire de la thyroïde a été décrite pour la première fois par Halsted en 1907. Dès les années 1930, la visualisation avec dissection systématique des nerfs récurrents durant l'opération était proposée pour en prévenir efficacement les lésions.^{161,162,163} Depuis ces découvertes, aucun changement notable n'est intervenu sur la technique chirurgicale malgré

certaines avancées liées à l'anesthésie ou aux équipements opératoires.¹⁶⁴ En 1966, l'ajout du neuromonitoring a été proposé pour suivre le trajet anatomique du nerf et ainsi en faciliter l'identification et la dissection.¹⁶⁵ Néanmoins, son intérêt pour prévenir la lésion du nerf demeure controversé,¹⁶⁶ au regard du manque de preuves disponibles dans la littérature objectivant son efficacité et de la nécessaire standardisation de ses modalités d'utilisation.¹⁶⁷ Seule une étude randomisée récente, menée au sein d'une équipe de trois chirurgiens, a suggéré son effet pour prévenir les paralysies récurrentielles transitoires.¹⁶⁸ D'autres voies d'amélioration sont également investiguées aujourd'hui, reposant sur l'anesthésie locale,¹⁶⁹ la chirurgie ambulatoire ou de court séjour,¹⁷⁰ la dissection par ultrasons,¹⁷¹ les techniques chirurgicales mini-invasives et endoscopiques¹⁷² ou le radioguidage pour traiter le cancer de la thyroïde.¹⁷³ En dehors de la technique opératoire elle-même, des précautions peuvent également être envisagées afin de prévenir la survenue d'une hypocalcémie au décours de l'intervention par l'administration périopératoire d'un traitement substitutif vitamino-calcique.¹⁷⁴

Dès 1931, Hegner décrivait la thyroïdectomie comme une chirurgie bien établie et relativement standardisée.¹⁷⁵ La clé d'une thyroïdectomie sûre et efficace repose sur le respect d'un plan opératoire rigoureux, progressant selon une série d'étapes logiques et ordonnées, basées sur une compréhension approfondie de l'anatomie et de la pathologie thyroïdienne.¹⁷⁶ Avec une technique appropriée, elle peut être effectuée avec un risque minimal pour les nerfs récurrents, les branches externes des nerfs laryngés supérieurs et les parathyroïdes. Tout d'abord, une exposition satisfaisante de la glande thyroïde est essentielle, le patient étant placé en position dorsale avec un support placé sous ses épaules pour permettre l'extension appropriée du cou. La dissection du pôle supérieur de la glande peut alors être réalisée, en utilisant le plan avasculaire situé entre le pôle et le muscle cricothyroïdien dans le but d'identifier et de préserver la branche externe du nerf laryngé supérieur. Par une rétraction sur la ligne médiane de la glande, on peut ensuite disséquer la face latérale du lobe de la thyroïde. En ayant recours à la technique de la dissection capsulaire, par ligature des branches de l'artère thyroïdienne inférieure, on s'efforce de préserver au mieux le nerf laryngé récurrent et les glandes parathyroïdes. Si l'on ne peut préserver une des parathyroïdes ou si l'une d'elles devient ischémique après la dissection de son pédicule vasculaire, son auto-transplantation doit être envisagée dans le muscle sternocléidomastoïdien homolatéral. Une fois l'hémostase réalisée de manière attentive, on procède alors à la fermeture.¹³⁷

2.3.3. Suivi des complications liées à la chirurgie thyroïdienne

La chirurgie de la thyroïde expose à la survenue de trois complications majeures, souvent réversibles et potentiellement évitables : la paralysie récurrentielle, l'hypoparathyroïdie secondaire et l'hématome compressif (table VIII).^{152,177,178} La survenue d'un hématome compressif du cou est rarissime (moins de 1%) et impose une réintervention en urgence compte tenu du risque vital immédiat causé par l'obstruction des voies respiratoires du patient. Une douleur postopératoire ou la cicatrice laissée au niveau de l'incision peuvent également incommoder le patient, d'autant plus en cas de complications de la plaie opératoire. Toutefois, la gêne occasionnée n'est que transitoire dans la grande majorité des cas.¹⁴⁰ A l'inverse, un abcès, des douleurs neuropathiques ou une lymphorrhée secondaires à un curage ganglionnaire étendu peuvent avoir des conséquences handicapantes pour le patient.

Table VIII. Complications de la thyroïdectomie

Paralysie récurrentielle partielle (la corde vocale bouge jusqu'à la ligne médiane) ou totale (la corde vocale est immobile), transitoire ou définitive
Hypoparathyroïdie secondaire transitoire ou définitive
Hémorragie avec risque d'hématome compressif en l'absence de prise en charge adaptée immédiate
Complications de la plaie opératoire liée à une infection, une cicatrice hypertrophique ou chéloïde
Obstruction des voies respiratoires par œdème laryngé, hématome compressif, paralysie récurrentielle bilatérale ou trachéomalacie
Douleur postopératoire
Abcès
Lymphorrhée
Pneumothorax
Embolie gazeuse
Crise aiguë thyrotoxique
Récidive de la pathologie
Hypothyroïdie

La paralysie récurrentielle, qui est provoquée par la lésion d'un ou deux nerf(s) récurrent(s) selon divers mécanismes liés à la technique du chirurgien,¹⁷⁹ peut entraîner une réduction partielle ou totale de la mobilité des cordes vocales. Habituellement bien tolérée, la paralysie unilatérale du nerf récurrent est néanmoins susceptible de provoquer une dysphonie

(voix bitonale, faible, rauque), une dysphagie (troubles de la déglutition accompagnés de fausses routes avec risque de pneumopathie d'inhalation) ou une dyspnée (gêne respiratoire à l'inspiration ou à l'effort). Elle peut altérer sérieusement la qualité de vie du patient^{180,181} et constitue la première cause de litige en chirurgie endocrinienne.^{182,183} Une rééducation de la voix peut accélérer sa récupération. La paralysie bilatérale est exceptionnelle mais gravissime car elle peut engager le pronostic vital du patient à court terme du fait de troubles respiratoires aigus. Elle peut avoir pour conséquence une réintubation pendant quelques jours, voire la confection d'une trachéotomie transitoire ou d'un geste ORL afin d'élargir l'espace entre les deux cordes vocales.¹⁴⁰ Parmi les principaux facteurs de risque connus de paralysie récurrentielle permanente figurent la réalisation d'une exérèse étendue de la glande thyroïde et le goitre récurrent.^{184,185}

L'hypoparathyroïdie secondaire est consécutive à la lésion d'au moins trois glandes parathyroïdes ou des vaisseaux les irriguant. Par conséquent, elle ne peut survenir qu'après une thyroïdectomie bilatérale et la préservation d'au moins deux glandes parathyroïdes, identifiées en cours d'opération, minimise significativement le risque d'hypoparathyroïdie.¹⁸⁶ Elle est responsable d'une hypocalcémie postopératoire qui se traduit cliniquement par des manifestations aiguës d'hyperexcitabilité neuromusculaire (paresthésies, hypoesthésies, contractures musculaires accompagnées de douleurs, d'angoisse, de polypnée, parfois de laryngospasme ou de convulsions, troubles cardiaques) et des manifestations chroniques (cataracte précoce, symptomatologie psychiatrique, ostéomalacie, troubles trophiques).¹⁸⁷ La prescription d'un traitement substitutif vitamino-calcique permet de contrôler ces manifestations. Parmi les principaux facteurs favorisant la survenue d'une hypoparathyroïdie postopératoire, on retiendra le sexe féminin, la maladie de Basedow, le goitre récurrent et l'étendue de la résection.^{184,186} Une chirurgie pour cancer de la thyroïde invasif ou métastatique augmente significativement le risque de ces deux complications¹⁸⁸ du fait de l'étendue de l'exérèse et du curage ganglionnaire qui peut y être associé.¹⁸⁹

La paralysie récurrentielle et l'hypoparathyroïdie doivent être recherchées systématiquement avant, puis au décours de toute intervention sur la thyroïde, afin de déterminer la responsabilité de l'acte dans leur survenue et plus particulièrement en cas de cancer infiltrant, de chirurgie thyroïdienne antérieure ou de signes cliniques évocateurs. Idéalement, leur caractère définitif doit être précisé à distance de l'acte. Toutefois, leurs modalités de détection sont encore mal codifiées. Nous avons identifié les recommandations de pratiques cliniques disponibles à ce jour sur le sujet, à partir d'une recherche Internet

systematique sur les sites d'organismes nationaux^{II} et de sociétés savantes^{III} responsables de leur développement en chirurgie endocrinienne, ainsi qu'à l'aide de méta-moteurs de recherche spécifiques,^{IV} de référentiels indexés dans Medline ou publiés dans des journaux spécialisés.^V Cette recherche montre qu'il n'existe pas de conduite à tenir consensuelle pour le suivi des complications de la chirurgie thyroïdienne, notamment en raison de l'absence de données scientifiques suffisantes. Seule la réalisation d'une laryngoscopie semble établie, sans précision sur son délai de réalisation après l'opération (Table IX).

Table IX. Principales recommandations disponibles pour le suivi des complications de la chirurgie thyroïdienne

Recommandation (Organisme, pays, date)	Niveaux de preuve scientifique	Hypoparathyroïdie secondaire	Paralysie récurrentielle
Paralysies récurrentielles de l'adulte (HAS, France, 2002) ¹⁹⁰	Non	-	Examen ORL avec laryngoscopie indirecte
Note d'information au patient (AFCE, France) ¹⁹¹	Non	-	Examen ORL avec laryngoscopie en cas de dysphonie postopératoire
Guidelines for the surgical management of endocrine disease and training requirements for endocrine surgery (BAES, Royaume-Uni, 2003) ¹⁹²	Non	Calcémie avant toute exploration du cou puis au décours d'une chirurgie bilatérale (< 24 heures et répété tant qu'elle ne s'est pas normalisée)	Laryngoscopie préopératoire en cas d'antécédent de dysphonie ou de cancer et postopératoire si dysphonie.
Guidelines for the management of thyroid cancer (BAES, Royaume-Uni, 2008) ¹⁹³	Oui	Surveillance de la calcémie à J1 puis chaque jour tant qu'elle n'est pas normalisée (III, B).	Laryngoscopie en cas de trouble de la voix persistant > 2 semaines après la chirurgie (IV, C).
Australian Endocrine Surgeons guidelines: Postoperative parathyroid hormone measurement and early discharge after total thyroidectomy (AES, Australie, 2007) ¹⁹⁴	Non	Parathormone 4 heures après la thyroïdectomie pour décider de la mise sous traitement substitutif puis suivi de la calcémie à distance pour adapter les posologies.	-

^{II} Haute Autorité de Santé, National Authority for Health, British National Institute of Clinical Excellence, American National Guidelines Clearinghouse, Guidelines International Network, Ontario Guidelines Advisory Committee.

^{III} Association Francophone de Chirurgie Endocrinienne, Société Française d'Endocrinologie, American Association of Endocrine Surgeons, American Thyroid Association, British Association of Endocrine Surgeons, European Society of Endocrine Surgeons, International Association of Endocrine Surgeons.

^{IV} Google Scholar, Tripdatabase, UpToDate, SumSearch.

^V Evidence Based Medicine et ACP journal club.

Nous avons poursuivi cette analyse bibliographique par une analyse systématique des revues de la littérature sur les méthodes de suivi des complications après chirurgie thyroïdienne. La recherche d'articles publiés, en langue française ou anglaise, au cours des vingt dernières années (jusqu'au mois de février 2009) a été menée à partir des bases suivantes : Cochrane Library, Medline et Web of Science. Deux algorithmes de recherche ont été conçus,¹⁹⁵ basés à la fois sur les mots clés de la base Medline (MeSH terms) et des mots clés libres recherchés dans le titre ou le résumé des papiers.^{VI} Seules les revues et les méta-analyses relatives au suivi des complications en chirurgie thyroïdienne ont finalement été retenues, en procédant successivement à la lecture de leur titre, de leur résumé et éventuellement de leur contenu. Deux relecteurs ont appliqué ces critères de sélection à l'aveugle et à tous les articles identifiés par les algorithmes. Des 659 références initialement identifiées par les algorithmes, nous en avons finalement retenu six, incluant quatre revues non systématiques^{196,197,198,199} et deux revues systématiques de la littérature.^{200,201} Après extraction de leur contenu à l'aide d'une grille standardisée, nous avons classé les méthodes de suivi des complications décrites en fonction de leur délai de réalisation après l'intervention, de leur performance diagnostique et de leur faisabilité en routine (table X).

Pour la détection d'une paralysie récurrentielle, cinq approches étaient proposées. La clinique présentait l'avantage d'être facilement applicable mais ne détectait pas de nombreux cas asymptomatiques (faible sensibilité). La laryngoscopie directe (au nasofibroscope) représentait l'examen de référence pour observer directement la mobilité des cordes vocales, malgré la détection possible de faux positifs lorsqu'elle était réalisée au décours immédiat de l'intervention (à l'extubation en salle de réveil). En cas de contraintes organisationnelles empêchant la réalisation systématique d'un tel examen en routine, la laryngoscopie indirecte (au miroir) devait être privilégiée bien qu'elle puisse être mal tolérée par le patient. L'électromyographie était intéressante pour affiner le diagnostic étiologique, en distinguant les troubles de mobilité du larynx d'origine neurogène des atteintes mécaniques liées à l'intubation. Cependant, elle était difficilement faisable en routine, nécessitait une grande

^{VI}Paralysies récurrentielle : ("Vocal Cord Paralysis"[Mesh] OR "Recurrent Laryngeal Nerve"[Mesh] OR "Voice Disorders"[Mesh] OR RLNP OR ("recurrent laryngeal nerve" OR "recurrent nerve" OR "vocal cord" OR "vocal fold")) AND (paralysis OR palsy OR injury OR lesion OR damage))) AND (Review[pt] OR meta-analysis* OR Cochrane).

Hypoparathyroïdie : ("Hypocalcemia"[Mesh] OR "Hypoparathyroidism"[Mesh]) AND (diagnosis OR screening OR "Biological Assay"[Mesh] OR "Laboratory Techniques and Procedures"[Mesh] OR "Diagnostic Techniques and Procedures"[Mesh] OR assay OR assessment) AND (Review[pt] OR meta-analysis* OR Cochrane).

expérience de l'opérateur pour sa réalisation et s'avérait particulièrement invasive pour le patient. La vidéostroboscopie était recommandée pour l'analyse fine des capacités vibratoires même si elle nécessitait un équipement spécialisé et des conditions de réalisation difficilement applicables en pratique courante. Actuellement à l'étude, l'échographie cervicale constitue un moyen non invasif et facile à mettre en œuvre pour visualiser la mobilité des cordes vocales. L'utilisation du neuromonitoring pour vérifier l'absence de lésion des nerfs récurrents en fin d'intervention est également étudiée. Bien que prometteurs, la performance diagnostique de ces deux examens doit encore être évaluée avant d'en envisager l'utilisation.

Concernant la détection des hypoparathyroïdies secondaires, quatre examens biologiques étaient décrits. L'analyse de la cinétique périopératoire de la parathormone permettait le dépistage précoce d'une anomalie, sa valeur pronostique d'une hypocalcémie à distance de l'opération étant particulièrement appréciable en cas de chirurgie ambulatoire ou de court séjour. Le dosage du calcium ionisé représentait la technique la plus performante mais ses conditions de prélèvement et de transport en limitaient fortement la faisabilité. A l'inverse, le dosage de la calcémie totale présentait l'avantage d'être facilement réalisable, au décours immédiat mais également à distance de l'acte. Ses performances pouvaient être améliorées par le calcul d'une calcémie corrigée, nécessitant toutefois le dosage et l'interprétation supplémentaires d'autres marqueurs biologiques.

Table X. Méthodes de détection des complications de la chirurgie thyroïdienne

	Test de détection	Performance	Faisabilité
Paralysie récurrentielle	Clinique	Basse	Élevée
	Echographie cervicale	Inconnue	Élevée
	Neuromonitoring	Inconnue	Moyenne
	Électromyographie	Élevée	Basse
	Vidéostroboscopie	Élevée	Basse
	Laryngoscopie indirecte	Moyenne	Moyenne
	Laryngoscopie directe	Élevée	Moyenne
Hypoparathyroïdie	Parathormone	Moyenne	Moyenne
	Calcium ionisé	Élevée	Basse
	Calcémie corrigée	Moyenne	Moyenne
	Calcémie totale	Moyenne	Élevée

Enfin, nous avons analysé les méthodes de détection utilisées dans le cadre de 4 grandes études multicentriques évaluant la survenue de complications après chirurgie thyroïdienne (table XI). L'examen de choix pour la détection des paralysies récurrentielles était la laryngoscopie, réalisée immédiatement après l'acte (complication transitoire) puis à 6 ou 12 mois (complication permanente). Le diagnostic d'hypoparathyroïdie secondaire reposait habituellement sur le dosage de la calcémie, de la parathormone ou sur l'administration d'un traitement substitutif vitamino-calcique au décours immédiat de l'acte (complication transitoire) puis à 6 ou 12 mois (complication permanente).

Table XI. Études multicentriques évaluant la survenue de complications après chirurgie thyroïdienne

Auteur, Année	Design	Pays (Période), Nombre de sujets	Taux de complications selon le protocole	
			Paralysie récurrentielle	Hypoparathyroïdie
Rosato, 2004 ²⁰²	Rétrospectif	Italie (1995-2000), 14 934 patients	T: 3,4% <i>Laryngoscopie</i> P (> M12): 1,0% <i>Laryngoscopie</i>	T: 10,0% <i>Calcémie</i> P (> M12): 1,7% <i>Calcémie</i>
Thomusch, 2000 ¹⁸⁴	Prospectif	Allemagne (1998), 7 266 patients Goitre bénin	T: 2,1% <i>Laryngoscopie</i> P (> M6): 1,1% <i>Laryngoscopie</i>	T: 6,4% <i>Substitution</i> P (> M6): 1,5% <i>Substitution</i>
Dralle, 2004 ¹⁸⁵	Prospectif	Allemagne (1998-2001), 16 448 interventions	P (> M6): 0,8% <i>Laryngoscopie</i>	
Thomusch, 2003 ¹⁸⁶	Prospectif	Allemagne (1998), 5 846 patients Chirurgie bilatérale		T: 7,3% <i>Clinique + substitution</i> P (> M6): 1,5% <i>PTH ou substitution</i>

T : temporaire ; P : permanente ; M : mois.

2.3.4. La mesure de la sécurité du patient en chirurgie thyroïdienne

La mesure de la sécurité du patient en chirurgie thyroïdienne repose principalement sur deux complications : la paralysie récurrentielle et l'hypoparathyroïdie secondaire. Les estimations réalisées sur leur survenue varient beaucoup selon les équipes et sont difficilement comparables pour des raisons méthodologiques (hétérogénéité des méthodes et délais de détection utilisés, diversité des pathologies traitées et des populations de patients recrutés, etc.). Les taux d'hypoparathyroïdies postopératoires rapportés vont de 1,6 à 53,6% pour les hypocalcémies transitoires et de 0 à 9,3% pour celles permanentes.²⁰³ Les taux moyens de paralysies récurrentielles postopératoires transitoires et permanentes sont estimés à 9,8% (allant de 0 à 38,4%) et 2,3% (allant de 0 à 18,6%), respectivement.²⁰⁴

La plupart des études portant sur la qualité des pratiques en chirurgie thyroïdienne ont surtout cherché à comparer la performance de chirurgiens expérimentés à celle de chirurgiens en cours de formation ou récemment formés, pour différentes indications opératoires (pathologie bénignes ou malignes) et différents types de chirurgie (unilatérale ou bilatérale). De qualité méthodologique parfois discutable, ces études ne sont pas parvenues à montrer de lien significatif entre expérience de l'opérateur et résultats de la chirurgie thyroïdienne.^{205,206,207,208,209}

Contrairement à d'autres spécialités, la relation entre volume d'activité et résultat de la prise en charge chirurgicale (*volume-outcome relationship*) a été peu explorée dans le cadre des pathologies thyroïdiennes. Aux États-Unis, des travaux menés à l'aide de bases de données hospitalières médico-administratives ont montré que le taux de complications après chirurgie de la thyroïde augmentait en cas de faible volume d'activité par chirurgien^{210,211} ou par hôpital.²¹² En Allemagne, l'équipe de Dralle a montré de manière prospective que ce lien existait spécifiquement entre le volume d'interventions par hôpital et la survenue d'une hypocalcémie¹⁸⁴ ou d'une paralysie récurrentielle¹⁸⁵ permanentes à 6 mois. Par contre, il n'existait pas de lien significatif entre le volume d'activité des chirurgiens et les résultats obtenus.¹⁸⁶

A notre connaissance, une seule étude a évalué la relation entre volume d'activité et qualité de la prise en charge chirurgicale (*volume-process relationship*) des patients atteints d'affections thyroïdiennes. Nous avons évalué, à l'aide de données hospitalières médico-administratives, l'adhésion des équipes chirurgicales à une indication opératoire bilatérale dans le cas du cancer de la thyroïde, en fonction de leurs volumes d'activité respectifs (toutes

pathologies thyroïdiennes confondues). Le risque de non-conformité à une chirurgie bilatérale d'emblée était plus important en cas de prise en charge dans un établissement à bas volume d'activité (*voir annexe 10.1. Lien entre volume d'activité et indication opératoire, article 5*).²¹³

3. Objectifs

Avant d'utiliser la carte de contrôle en routine, il convenait de s'assurer de sa transposabilité au domaine de la sécurité du patient en chirurgie, ce qui comprenait à la fois :

- son adaptabilité à la complexité et à la variabilité des données sur les caractéristiques cliniques du patient ;
- son acceptabilité par les professionnels de santé et son applicabilité sur le terrain face à des contraintes opérationnelles ;
- et finalement son utilité, aussi bien pour un contrôle en routine et une amélioration continue de la performance des chirurgiens au cours du temps, que pour investiguer ponctuellement l'impact d'une action concrète visant à améliorer la sécurité du patient opéré.

L'objectif de ce travail était d'évaluer l'intérêt d'un suivi d'indicateurs par cartes de contrôle lorsqu'il était appliqué aux résultats de la chirurgie thyroïdienne. Cet objectif global était déclinable en trois sous-objectifs :

3.1. Évaluer la faisabilité et l'utilité d'un suivi d'indicateurs par cartes de contrôle dans la pratique courante d'une équipe chirurgicale

3.2. Développer une carte de contrôle adaptée au suivi des complications de la chirurgie thyroïdienne

3.2.1. Comparaison de performance entre une carte à limites approximées et une carte à limites exactes

3.2.2. Comparaison de performance entre une carte à limites brutes et une carte à limites ajustées sur les caractéristiques des patients

3.3. Étudier l'effet du neuromonitoring sur la performance individuelle du chirurgien en traçant la courbe d'apprentissage liée à son utilisation

4. Approche commune à l'ensemble des travaux

4.1. Population étudiée et sources d'informations

Cette étude a consisté en une observation longitudinale de l'activité de chirurgie thyroïdienne réalisée dans le service de chirurgie générale et endocrinienne du Centre Hospitalier Lyon Sud (Hospices Civils de Lyon). Elle s'est déroulée sur quatre années, de 2006 à 2009. Chaque année, plus de 700 interventions sur la thyroïde étaient pratiquées par une équipe de trois chirurgiens endocriniens, incluant un chirurgien expérimenté à très haut volume d'activité (plus de 300 interventions sur la thyroïde par an depuis 25 ans) et deux jeunes chirurgiens à haut volume d'activité (chacun réalisant plus de 150 interventions sur la thyroïde par an depuis 5 ans).

4.1.1. Technique opératoire

La technique chirurgicale couramment utilisée par les chirurgiens du service reposait sur un plan opératoire rigoureux, progressant selon une série d'étapes prédéterminées. Après une incision de Kocher, un rabat des tissus musculo-cutanés était réalisé sans section des muscles, lesquels étaient séparés au niveau de la ligne médiane pour permettre l'exposition de la glande thyroïde. Ensuite, un des lobes thyroïdiens était extériorisé et la dissection commençait par la ligature des vaisseaux thyroïdiens supérieurs. Durant cette étape, des précautions étaient prises pour éviter de léser la branche externe du nerf laryngé supérieur. Puis le nerf récurrent était systématiquement visualisé en utilisant l'artère thyroïdienne inférieure comme point de repère. Au cours de la dissection du nerf, le chirurgien s'efforçait de conserver une vascularisation satisfaisante des glandes parathyroïdes en ligaturant les vaisseaux thyroïdiens proches de la capsule thyroïdienne. En cas de thyroïdectomie totale, le lobe controlatéral était réséqué de la même manière.

Le monitoring du nerf récurrent n'a été mis en place dans le service qu'à partir de mai 2008 et son utilisation était laissée à l'initiative du chirurgien. Durant l'induction, si le chirurgien avait décidé d'utiliser le neuromonitoring, l'anesthésiste procédait alors à l'installation d'électrodes endotrachéales avant l'intubation. Cette technique autorisait une méthode d'enregistrement stable et continu de l'électromyogramme laryngé. Les cordes vocales étaient systématiquement visualisées de telle sorte qu'elles soient en contact avec les électrodes endotrachéales gauche et droite. Le *NIM-Response Nerve Integrity Monitoring System* (Xomed Medtronics, Jacksonville, FL, USA) a été utilisé avec une amplitude de

stimulation maximale fixée à 2 mA (figure 11). Le neuromonitoring était utilisé à la fois pour confirmer le trajet du nerf après son identification ou bien pour aider à le localiser en cas de visualisation difficile. Le nerf récurrent était alors stimulé en appliquant la sonde du neuromonitoring directement sur le nerf (si visualisé) ou sur l'aire de dissection. Un signal positif était recherché par l'observation de l'électromyogramme. Le nerf vague était stimulé entre l'artère carotide et la veine jugulaire interne en cas de perte de signal au moment où survenait une lésion, afin de différencier un éventuel artefact d'une réelle dysfonction du nerf. De même, le nerf était stimulé à la fin de l'exérèse, dans le but de s'assurer de l'absence de lésion récurrentielle.

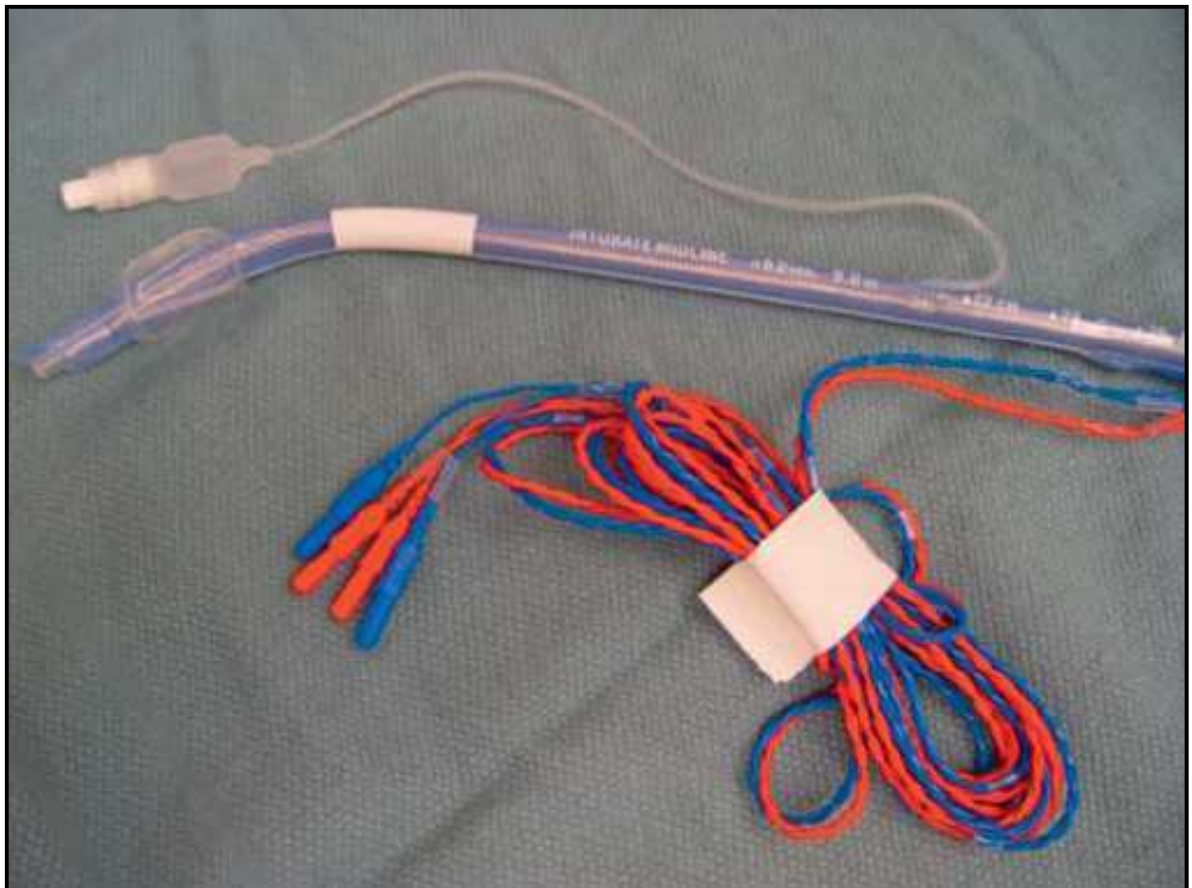


Figure 11. Photographie des composants essentiels du neuromonitoring

4.1.2. Design d'étude

L'évaluation a été conduite selon deux temps successifs (figure 12) :

1) Un état des lieux initial a été réalisé avant la mise en place des cartes de contrôle. Afin d'obtenir une estimation de la proportion de complications à la fois précise (distance par rapport à la valeur observée de 5%) et représentative de l'activité du service, un échantillon de 346 patients opérés de la thyroïde entre le 1^{er} janvier 2004 et le 31 décembre 2005 a été sélectionné par tirage au sort avec stratification sur le mois de l'intervention. Les dossiers médicaux de ces patients ont été audités et des données relatives aux caractéristiques des patients, aux types de chirurgie réalisée et aux complications survenues ont été collectées rétrospectivement au cours du premier trimestre 2006. Ces informations devaient nous permettre de paramétrer les cartes de contrôle et de comparer la performance des chirurgiens avant et après leur instauration dans le service.

2) Un suivi prospectif continu a débuté dès la mise en place des cartes de contrôle. Tous les patients ayant bénéficié d'une intervention sur la thyroïde entre le 1^{er} janvier 2006 et le 30 avril 2009 étaient incluables. Au total, 2 115 patients sur les 2 438 opérés durant ces 40 mois de suivi ont été inclus (exhaustivité des inclusions de 86,8%).^{VII} L'âge médian des patients opérés était de 52,8 ans (étendue de 8,8 à 93,5 ans) et 78,6% étaient des femmes. La pathologie thyroïdienne sous-jacente était habituellement un goitre (63,3% des cas), un cancer (23,6%) ou un nodule (12%). L'acte réalisé sur la thyroïde était unilatéral dans 12,3% des cas (lombo-isthmectomie) alors qu'il s'agissait d'une thyroïdectomie totale ou subtotale dans 73,9% des cas, d'une chirurgie élargie avec curage dans 7% des cas ou d'une totalisation secondaire dans 5,9% des cas. La durée moyenne de séjour était de 2,8 jours (écart-type : 1,3).

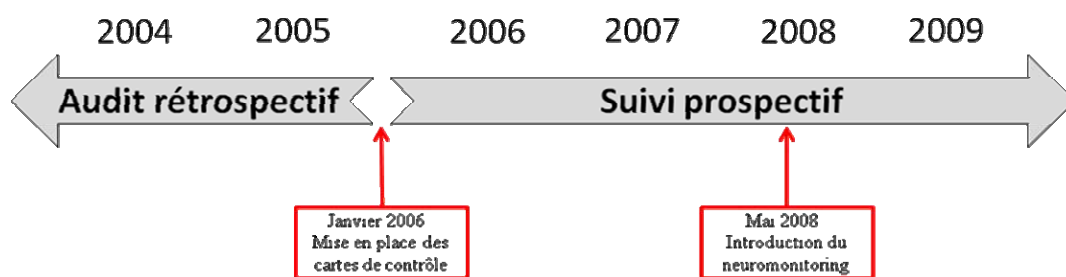


Figure 12. Schéma du design longitudinal de l'étude

^{VII} Parmi les 323 patients non inclus l'âge médian était de 54,7 ans et 64,7% étaient des femmes. La prise en charge était liée spécifiquement à une intervention sur la thyroïde dans 43% des cas (contre 85% des cas parmi les patients inclus), à une intervention au cours de lymphome ou de leucémie dans 18,3% des cas ou à une intervention sur les parathyroïdes dans 7,3% des cas.

4.1.3. Protocole d'évaluation

Au cours du suivi prospectif, un protocole de détection des complications et de collecte de l'information a été mis en œuvre spécifiquement (figure 13). A l'issue de l'analyse de la littérature présentée précédemment (*voir partie 2.3.3. Suivi des complications opératoires liées à la chirurgie thyroïdienne*), il nous a semblé que le meilleur équilibre entre performance et faisabilité de la méthode de détection pour suivre les complications immédiates de la chirurgie thyroïdienne reposait sur la laryngoscopie directe et le dosage de la calcémie totale dans les 48 heures suivant l'intervention. Après réalisation d'une chirurgie thyroïdienne unilatérale ou bilatérale par le chirurgien, la mobilité des cordes vocales du patient était donc systématiquement vérifiée par l'anesthésiste au moment de l'extubation à l'aide d'un nasofibroscope. Au cas où ce premier dépistage s'avérait non concluant ou en cas de symptômes évocateurs d'une lésion récurrentielle (voix anormale, troubles de la déglutition ou respiratoires), l'examen était renouvelé par un médecin ORL dans les 48 heures suivant l'intervention. Parallèlement, un dosage de la calcémie totale était systématiquement réalisé le jour suivant l'intervention par le biologiste, et répété en cas de résultat anormal lors du premier dosage (calcémie < 2 mmol/L ou 8,0 mg/dL) afin de détecter la survenue d'une hypoparathyroïdie secondaire. La valeur normale du laboratoire pour la calcémie totale était comprise entre 2,25 et 2,60 mmol/L. Aucun supplément vitamino-calcique n'était administré par voie orale au patient avant la chirurgie mais du calcium et/ou de la vitamine D pouvaient lui être prescrits si nécessaire après la chirurgie.

A la sortie du patient, des informations sur son hospitalisation étaient systématiquement recueillies par l'interne au moyen d'une fiche de codage conservée dans le dossier médical (*voir annexe 10.2. Fiche de codage*). Ces informations étaient ensuite saisies par la secrétaire dans la base du Programme de Médicalisation des Systèmes d'Information (PMSI) des Hospices Civils de Lyon enrichie de codes supplémentaires. Elles comprenaient principalement les caractéristiques démographiques du patient (âge, sexe), la pathologie thyroïdienne prise en charge, la nature de l'acte chirurgical réalisé et l'identité du premier opérateur, la survenue éventuelle d'une paralysie récurrentielle et/ou d'une hypocalcémie postopératoire et la durée du séjour.

Chaque trimestre, ces données étaient extraites de la base médico-administrative du PMSI par le méthodologiste pour permettre la construction des cartes de contrôle, lesquelles étaient ensuite discutées lors d'une réunion de restitution pluridisciplinaire rassemblant l'ensemble des professionnels impliqués dans le projet (chirurgiens, épidémiologistes,

anesthésistes, cadres de bloc, internes, secrétaires). Afin de faciliter l'interprétation des variations observées sur les indicateurs, nous nous sommes principalement aidés des éléments retranscrits dans le journal de bord tenu par les chirurgiens du service. Ce journal de bord répertoriait en temps réel et de manière exhaustive l'ensemble des changements (techniques, matériels, humains, organisationnels, etc.) intervenus sur la prise en charge standard des patients et ayant pu en influencer le résultat. Au cours de ces réunions, si une cause spéciale de variation était détectée par les cartes, nous nous sommes efforcés de l'identifier en vue de mettre en œuvre rapidement des mesures correctives pour en éviter la récurrence. Ces réunions étaient également l'occasion de débats orientant la conception d'actions d'amélioration de la prise en charge à mettre en œuvre puis à tester.

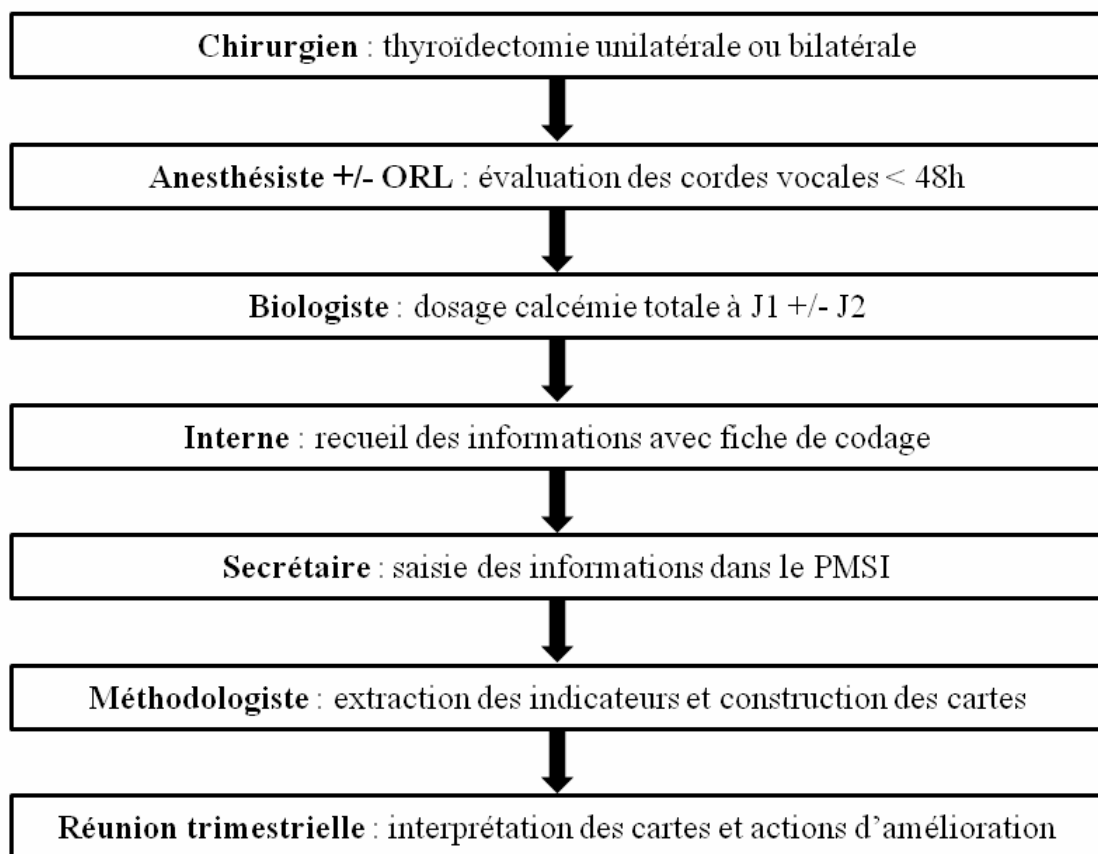


Figure 13. Étapes successives du protocole d'évaluation

4.2. Indicateurs d'évaluation

Deux indicateurs ont été utilisés pour mesurer la sécurité de la chirurgie thyroïdienne en fonction de ses principales complications :

- la paralysie récurrentielle, dont la survenue était évaluée par la détection d'une altération de la mobilité des cordes vocales postopératoire ;
- l'hypoparathyroïdie secondaire, dont la survenue était évaluée par la détection d'une hypocalcémie postopératoire.

Ces indicateurs ont d'abord été mesurés à partir des données recueillies lors de l'audit rétrospectif. Puis, au cours du suivi prospectif, ils étaient extraits de la base PMSI de manière quasi-automatisée en utilisant le logiciel *SAS version 9.13* (*SAS Institute, Cary, North Carolina, USA*).

4.2.1. Paralyse récurrentielle

Le premier indicateur était la proportion intra-hospitalière de patients ayant présenté une altération unilatérale ou bilatérale de la mobilité des cordes vocales immédiatement après l'intervention parmi l'ensemble des patients ayant bénéficié d'une chirurgie thyroïdienne unilatérale ou bilatérale (table XII). Les patients présentant une paralysie récurrentielle connue, préexistante à l'intervention, étaient identifiés comme tels et ne devaient donc pas être inclus dans l'analyse. Le caractère unilatéral ou bilatéral de la paralysie récurrentielle n'étant pas précisé, la mesure de la complication portait donc sur le nombre de patients opérés et non sur le nombre de nerfs à risques exposés à la chirurgie.

Table XII. Carte d'identité de l'indicateur estimant la survenue d'une paralysie récurrentielle

Domaine	Sécurité du patient en chirurgie thyroïdienne
Définition	Survenue d'une altération unilatérale ou bilatérale de la mobilité des cordes vocales dépistée au nasofibroscope dans les 48 heures suivant une intervention sur la thyroïde
Source d'information	PMSI avec renseignement d'un code interne supplémentaire
Population source	<u>Critères d'inclusion</u> : patient hospitalisé dans le service de chirurgie générale et endocrinienne du Centre Hospitalier Lyon Sud pour y bénéficier d'un acte chirurgical sur la thyroïde unilatéral ou bilatéral*
	<u>Critères d'exclusion</u> : patient présentant une paralysie récurrentielle antérieure à l'intervention, erreur de codage ou donnée manquante
Unité statistique	Séjour hospitalier avec réalisation d'un acte de chirurgie thyroïdienne unilatéral ou bilatéral
Méthode de calcul	<u>Numérateur</u> : nombre de séjours réalisés avec survenue d'une altération de la mobilité des cordes vocales en postopératoire immédiat pendant une période donnée
	<u>Dénominateur</u> : nombre de séjours réalisés pendant une période donnée
Comparaison	Temporelle avec suivi des variations au cours du temps
Fréquence de mesure	Mensuelle
Pertinence clinique	La paralysie récurrentielle constitue une complication majeure et la principale cause de litige au décours d'une chirurgie thyroïdienne par lésion du(des) nerf(s) récurrent(s). La nasofibroscopie avec visualisation directe de la mobilité des cordes vocales représente l'examen de référence pour en détecter la survenue.

* La détection des séjours avec chirurgie thyroïdienne unilatérale ou bilatérale reposait sur les codes de la Classification Commune des Actes Médicaux suivants : KCFA001, KCFA008, KCFA010, KCFA009, KCFA005, KCFA007, KCFA002, KCFA003, KCFA006, KCMA001, KCJB001, KCJA001, KCGA001, KCFA004.

4.2.2. Hypoparathyroïdie secondaire

Le second indicateur était la proportion intra-hospitalière de patients ayant présenté une hypocalcémie postopératoire immédiatement après l'intervention parmi l'ensemble des patients ayant bénéficié d'une chirurgie thyroïdienne bilatérale (table XIII). L'hypocalcémie était déterminée selon un seuil de détection fixé à 2 mmol/L. Seule la valeur la plus basse de l'ensemble des dosages de calcémie totale réalisés au cours du séjour d'un même patient était retenue pour la mesure de l'indicateur. Les patients présentant une hypoparathyroïdie connue, préexistante à l'intervention, étaient identifiés comme tels et ne devaient donc pas être inclus dans l'analyse.

Table XIII. Carte d'identité de l'indicateur estimant la survenue d'une hypoparathyroïdie secondaire

Domaine	Sécurité du patient en chirurgie thyroïdienne
Définition	Survenue d'une hypocalcémie dépistée par au moins un dosage de calcémie totale < 2 mmol/L dans les 48 heures suivant une intervention sur la thyroïde
Source d'information	PMSI avec renseignement d'un code interne supplémentaire
Population source	<p><u>Critères d'inclusion</u> : patient hospitalisé dans le service de chirurgie générale et endocrinienne du Centre Hospitalier Lyon Sud pour y bénéficier d'un acte chirurgical sur la thyroïde bilatéral*</p> <p><u>Critères d'exclusion</u> : patient présentant une hypoparathyroïdie antérieure à l'intervention, erreur de codage ou donnée manquante</p>
Unité statistique	Séjour hospitalier avec réalisation d'un acte de chirurgie thyroïdienne bilatéral
Méthode de calcul	<p><u>Numérateur</u> : nombre de séjours réalisés avec survenue d'une hypocalcémie postopératoire immédiate pendant une période donnée</p> <p><u>Dénominateur</u> : nombre de séjours réalisés pendant une période donnée</p>
Comparaison	Temporelle avec suivi des variations au cours du temps
Fréquence de mesure	Mensuelle
Pertinence clinique	L'hypoparathyroïdie secondaire constitue une complication majeure au décours d'une chirurgie thyroïdienne par lésion d'au moins trois glandes parathyroïdes ou de leur vascularisation. En l'absence d'examen de référence, le dosage de la calcémie totale représente l'examen biologique le plus simple à mettre en œuvre pour détecter sa survenue.

* La détection des séjours avec chirurgie thyroïdienne bilatérale reposait sur les codes de la Classification Commune des Actes Médicaux suivants : KCFA010, KCFA009, KCFA005, KCFA007, KCFA002, KCFA003, KCFA006, KCMA001, KCJB001, KCJA001, KCGA001, KCFA004.

4.2.3. Méthodes d'ajustement

Afin de tenir compte des variations de recrutement au cours du temps, liées aux caractéristiques des patients opérés ou de leur prise en charge et susceptibles d'influencer la survenue de complications opératoires, deux modèles d'ajustement sur le *case-mix* ont été développés.

Nous avons retenu des variables cliniques documentées dans la littérature comme étant associées au résultat de la chirurgie, à savoir le sexe et l'âge du patient, ainsi que la pathologie thyroïdienne opérée et le type d'acte chirurgical réalisé.^{184,185,186} Des odds-ratios (OR) ajustés sur le *case-mix* ont été calculés pour chacune des complications opératoires étudiées (la paralysie récurrentielle et l'hypoparathyroïdie), au moyen de deux modèles de régression logistique appliqués à l'ensemble des données collectées de janvier 2006 à mai 2008. Pour chacune des chirurgies thyroïdiennes réalisées, les probabilités de survenue d'une paralysie récurrentielle et d'une hypoparathyroïdie secondaire ont été déterminées à partir des estimations du modèle, en fonction du risque préopératoire individuel lié à chaque patient. La proportion attendue de chacune des complications a ainsi été obtenue mensuellement en sommant leurs probabilités de survenue pour l'ensemble des patients opérés.

La performance de chaque modèle a été évaluée en fonction de deux paramètres. Nous avons d'abord mesuré le pouvoir discriminatoire du modèle à l'aide de la *c*-statistics (*Receiver Operating Characteristic curve score*). Une valeur d'aire sous la courbe ROC inférieure à 0,7 montrait la faible discrimination du modèle.²¹⁴ Puis, nous avons estimé l'adéquation du modèle aux données (*goodness of fit*) au moyen du test de Hosmer et Lemeshow. Une non-significativité du test ($p > 0,05$ puis valeur la plus proche possible de 1) était considérée comme satisfaisante.²¹⁵ D'éventuelles interactions entre les variables entrées dans les modèles étaient systématiquement recherchées.

L'ajustement sur le *case-mix* a été réalisé à l'aide du logiciel *SPSS version 12.0* (*SPSS Inc.*, Chicago, IL, USA).

4.3. Cartes de contrôle développées

4.3.1. Cartes de Shewhart

Afin de suivre la performance de l'équipe chirurgicale au cours du temps, chacun des indicateurs a été représenté sur une carte de contrôle longitudinale de type p .⁹¹ Chaque point représenté sur le graphique exprimait la proportion mensuelle de complications observées parmi les patients opérés. Les limites de contrôle et d'alerte ont été positionnées à une distance de 2 et 3 déviations standards (DS) autour de la ligne centrale, de telle sorte que la détection d'une cause spéciale était définie par : i) la présence d'un point en dehors des limites de contrôle inférieure ou supérieure ; ii) ou la présence de deux points sur trois consécutifs en dehors des limites d'alerte inférieure ou supérieure, d'un même côté de la ligne centrale.⁹⁹

Conception de cartes à limites brutes ou ajustées

Deux types de cartes ont été construits à l'aide du logiciel *Microsoft Office Excel 2007* (*Microsoft Corporation*, Redmond, WA, USA) :

- Une carte de contrôle à limites brutes : La valeur de ligne centrale était constante et correspondait à la proportion globale de complications observées sur l'ensemble de la période de suivi. Les limites inférieures et supérieures variaient mensuellement en fonction de la taille des échantillons de patients opérés chaque mois.
- Une carte de contrôle à limites ajustées sur le *case-mix* : La valeur de ligne centrale variait mensuellement et indiquait la proportion attendue de complications en fonction des caractéristiques des patients opérés chaque mois. Les limites inférieures et supérieures variaient mensuellement en fonction de la proportion attendue de complications et de la taille des échantillons de patients opérés chaque mois.

Calcul de limites approximées ou exactes

Deux méthodes de calcul ont été envisagées afin d'obtenir un tracé des limites approximées ou exactes sur les cartes. Soit i le nombre d'échantillons de taille variable n_i et comprenant un nombre variable d'évènements x_i . Soit p , la valeur de la ligne centrale correspondant à la proportion globale de complications observées sur l'ensemble de la période de suivi (\bar{p} pour une carte brute) ou bien à la proportion mensuelle attendue de complications

(p_i pour une carte ajustée). Classiquement, le calcul des limites d'une carte de contrôle de type p reposait sur une approximation normale de la distribution binomiale. En conséquence, des limites approximées ont d'abord été calculées en appliquant les formules suivantes :

- Limites de contrôle approximées $= p \pm 3\sqrt{\frac{p \times (1-p)}{n_i}}$;
- Limites d'alerte approximées $= p \pm 2\sqrt{\frac{p \times (1-p)}{n_i}}$.

Cette approximation était applicable tant que $n_i \bar{p}(1-\bar{p}) > 5$ et $0,1 \leq \bar{p} \leq 0,9$. Si ces conditions n'étaient pas satisfaites, il convenait de calculer des limites exactes reposant directement sur la loi de distribution de probabilité binomiale.⁹⁷ Conformément à la distribution normale, 95,45% de l'ensemble des points représentés sur la carte étaient supposés se trouver à une distance inférieure à 2 déviations standards autour de la ligne centrale, les 4,55% restants se situant au-delà. De même, 99,73% de l'ensemble des points étaient supposés être positionnés à une distance inférieure à 3 déviations standards autour de la ligne centrale si le processus était stable (causes communes de variation), seuls 0,27% d'entre eux dépassant cette limite (causes spéciales de variation).

Par conséquent, nous avons utilisé la distribution binomiale cumulée ($F(x_t, \bar{p}, n_t) = P(X_t \leq x_t) = \sum_{i=0}^{x_t} \binom{n_t}{i} \bar{p}^i (1-\bar{p})^{n_t-i}$) pour calculer les limites exactes à l'aide des formules suivantes :

- Limites de contrôle exactes pour 3 DS (99,73%)
 - inférieure : valeur de x_i qui satisfait $P(X_i \leq x_i) = \frac{1-0,9973}{2} = 0,00135$;
 - supérieure : valeur de x_i qui satisfait $P(X_i \leq x_i) = \frac{1+0,9973}{2} = 0,99865$;
- Limites d'alerte exactes pour 2 DS (95,45%)
 - inférieure : valeur de x_i qui satisfait $P(X_i \leq x_i) = \frac{1-0,9545}{2} = 0,02275$;
 - supérieure : valeur de x_i qui satisfait $P(X_i \leq x_i) = \frac{1+0,9545}{2} = 0,97725$.

4.3.2. Cartes CUSUM

Afin de suivre l'apprentissage du neuromonitoring par chaque chirurgien, des cartes de séries cumulées (cartes CUSUM) ont été élaborées.¹²⁸ La maîtrise de cet outil a été évaluée en représentant la survenue cumulée des succès (absence de complication) et des échecs (paralysie récurrentielle postopératoire) obtenus par chacun d'entre eux. Les courbes d'apprentissage correspondantes ont été tracées en utilisant deux modes de représentation graphique complémentaires, à l'aide du logiciel *Microsoft Office Excel 2007* (*Microsoft Corporation*, Redmond, WA, USA) :

- la carte cumulée d'échecs observés moins attendus (*cumulative observed minus expected failure chart* ou carte O-E)¹²² qui permettait une interprétation visuelle intuitive de la courbe d'apprentissage ;
- le test de séries cumulées lié à la courbe d'apprentissage (*cumulative summation test for learning curve* ou test LC-CUSUM)¹³¹ qui fournissait un test statistique afin de détecter formellement quand la nouvelle compétence était acquise par le chirurgien.

Carte O-E

La carte *O-E* montrait si la performance d'un chirurgien se situait au-delà ou en deçà de ce qui pouvait être espéré en représentant chronologiquement la différence entre le nombre observé (*observed*) et le nombre attendu (*expected*) de complications, opération après opération. Pour chaque opération, la probabilité de survenue d'une complication a été calculée à l'aide d'un modèle d'ajustement sur le *case-mix*, afin de déterminer l'amplitude avec laquelle le graphique montait ou descendait. Pour un succès, le graphique descendait d'une quantité égale à la probabilité estimée de survenue d'une complication. Pour un échec, le graphique montait d'une quantité égale à la probabilité estimée d'absence de survenue d'une complication. Démarrant à la valeur 0 (axe horizontal), le graphique tendait vers le haut si la proportion d'échecs cumulés augmentait au-delà de ce qui était attendu, alors qu'il tendait vers le bas si la proportion d'échecs cumulés diminuait en deçà de ce qui était attendu. Si la performance du chirurgien était acceptable en regard de ce qui avait été prédit par le modèle d'ajustement, le graphique oscillait autour de l'axe horizontal.

Test LC-CUSUM

Un test LC-CUSUM a été développé spécifiquement pour déterminer le nombre d'opérations nécessaires à l'apprentissage du neuromonitoring par chacun des chirurgiens. Considérant que son utilisation n'était pas contrôlée au début de l'apprentissage, une limite de décision inférieure h était figurée horizontalement pour détecter à partir de quand le chirurgien en avait acquis la maîtrise. Cette limite permettait de tester l'hypothèse nulle H_0 , correspondant à un niveau de performance jugé inacceptable (proportion de complications inacceptable p_0 et utilisation du neuromonitoring hors contrôle), contre l'hypothèse alternative H_1 , correspondant à un niveau de performance à atteindre (proportion de complications acceptable p_1 et utilisation du neuromonitoring sous contrôle). Nous avons fixé la valeur de p_1 en fonction de la prévalence de paralysies récurrentielles postopératoires observée avant l'introduction du neuromonitoring. Pour calculer p_0 , nous avons appliqué à p_1 une augmentation du risque de complications équivalente à un OR de 2. Ainsi, la réduction du risque de complications entre p_0 (14%) et p_1 (7,5%) pour détecter le passage d'une situation d'apprentissage à un niveau de performance acceptable était équivalente à un OR de 0,5. Graphiquement, l'utilisation du neuromonitoring était considérée non maîtrisée tant que la courbe d'apprentissage demeurait au-dessus de la limite h . Lorsque cette limite était franchie, suffisamment de preuves avaient été accumulées pour considérer que l'apprentissage du chirurgien était terminé.

Le score LC-CUSUM était représenté chronologiquement sur le graphique en fonction du nombre d'opérations réalisées au cours du temps t et son calcul reposait sur $S_t = \min(0, S_{t-1} - W_t)$, avec une valeur initiale de $S_0=0$ et W_t qui mesurait le poids de chaque opération supplémentaire. Considérant que le risque de complications pouvait varier entre les patients opérés, nous avons calculé p_t à partir d'un modèle d'ajustement sur le *case-mix* fournissant pour chaque opération une probabilité de survenue de complications X_t . Les hypothèses nulle et alternative, H_0 et H_1 , étaient exprimées en terme d'OR, telles que $H_0: \theta=\theta_0=1$ et $H_1: \theta=\theta_1=0.5$. Le score LC-CUSUM ajusté pour un résultat binaire (succès/échec de l'opération) était calculé en attribuant un poids par opération de :

$$\begin{aligned}
 - \quad W_t &= \log \left[\frac{1 - p_t + \theta_0 p_t}{1 - p_t + \theta_1 p_t} \right] = \log \left[\frac{1}{1 - 0.5 p_t} \right] \text{ si } X_t=0 \text{ (succès)} ; \\
 - \quad W_t &= \log \left[\frac{(1 - p_t + \theta_0 p_t) \theta_1}{(1 - p_t + \theta_1 p_t) \theta_0} \right] = \log \left[\frac{0.5}{1 - 0.5 p_t} \right] \text{ si } X_t=1 \text{ (échec)}.
 \end{aligned}$$

Le choix de la limite de décision h dépendait à la fois de p_1 et des performances du test LC-CUSUM qui étaient exprimées en termes d'*average run length* (ARL).¹³⁰ Les ARL étaient définis comme le nombre moyen d'opérations à réaliser avant qu'un signal ne survienne sous l'hypothèse nulle (ARL₀ est lié au taux de faux positifs) et l'hypothèse alternative (ARL₁ est lié au taux de vrais positifs). Par conséquent, h était déterminé arbitrairement dans le but d'atteindre un équilibre optimal maximisant ARL₀ et minimisant ARL₁. Différentes valeurs d'ARL ont été calculées à partir d'une simulation sur 10 000 échantillons réalisées à l'aide du logiciel *R* (*R Foundation for Statistical Computing*, Vienna, Austria). Nous avons finalement fixé h à -1,35 pour atteindre des valeurs d'ARL₀ et d'ARL₁ de 119 et 34, respectivement (table XIV). Concrètement, si l'utilisation du neuromonitoring était hors contrôle (apprentissage en cours), une fausse alarme était susceptible d'apparaître toutes les 119 opérations. En revanche, si l'utilisation du neuromonitoring était sous contrôle (apprentissage terminé), 34 opérations étaient nécessaires pour détecter un changement de performance.

Table XIV. Valeurs d'ARL₀ et d'ARL₁ du test LC-CUSUM en fonction de la limite de décision h choisie

h	ARL ₀	ARL ₁
- 1,00	59,3	21,4
- 1,05	63,8	23,0
- 1,10	73,7	25,2
- 1,15	75,7	26,0
- 1,20	88,7	28,2
- 1,25	104,9	31,8
- 1,30	113,8	33,2
- 1,35	118,8	33,7
- 1,40	128,0	35,3
- 1,45	139,3	37,2
- 1,50	152,5	38,6

5. Travaux scientifiques répondant aux objectifs

5.1. Étude d'applicabilité de la carte de Shewhart au contrôle qualité des résultats de la chirurgie thyroïdienne (*article 1*)

Lors de cette première partie du projet, nous avons expérimenté la mise en place de cartes de contrôle afin de suivre la performance collective d'une équipe de chirurgiens endocriniens. L'objectif de ce travail était de tester la faisabilité et l'utilité de la carte de Shewhart lors de son application routinière au contrôle qualité des résultats de la chirurgie thyroïdienne. Pour y répondre, deux approches méthodologiques ont été proposées.

Nous avons d'abord comparé les résultats de la chirurgie thyroïdienne avant (données d'audit rétrospectif sur 2004-2005) et après (données de suivi prospectif sur 2006-2007) l'introduction des cartes de contrôle. Une diminution importante des hypocalcémies postopératoires a été observée, passant de 32,3% à 20,9% des thyroïdectomies réalisées.

Nous avons ensuite suivi l'évolution mensuelle des indicateurs sur les cartes afin de détecter puis d'identifier d'éventuelles causes spéciales de variation. Au total, 1 114 interventions sur la thyroïde ont été évaluées de janvier 2006 à décembre 2007. Une dégradation significative de la performance des chirurgiens a été détectée au cours de l'été 2007. Deux causes spéciales concourant à la suractivité de l'équipe chirurgicale ont été identifiées :

- La rénovation du bloc opératoire habituellement dédié à la chirurgie thyroïdienne ayant conduit à un changement de salle temporaire avec réduction du nombre de plages opératoires disponibles en dépit du maintien d'un volume élevé de thyroïdectomies ;
- L'absence occasionnelle du chirurgien responsable de la majorité des interventions sur la thyroïde réalisées par l'équipe.

Antoine Duclos, Sandrine Touzet, Pietro Soardo, Cyrille Colin, Jean-Louis Peix, Jean-Christophe Lifante.

Quality monitoring in thyroid surgery using the Shewhart control chart.

British Journal of Surgery 2009;96(2):171-174.

Le premier article publié et reproduit ci-après présente l'expérience de la mise en œuvre de la carte de contrôle dans la pratique courante d'une équipe de chirurgiens. Il révèle l'intérêt de la méthode pour mieux comprendre les facteurs influençant le résultat d'une chirurgie et en définitive améliorer la sécurité du patient opéré.

Quality monitoring in thyroid surgery using the Shewhart control chart

A. Duclos¹, S. Touzet¹, P. Soardo², C. Colin¹, J. L. Peix² and J. C. Lifante²

¹Pôle d'Information Médicale Evaluation Recherche, Hospices Civils de Lyon, and Université Lyon 1, Hôtel Dieu, Lyon, and ²Hospices Civils de Lyon, Department of General and Endocrine Surgery, Centre Hospitalier Lyon Sud, Pierre Bénite, France

Correspondence to: Dr A. Duclos, Hospices Civils de Lyon, Pôle d'Information Médicale Evaluation Recherche, Unité d'Etude des Pratiques Professionnelles, 162 Avenue Lacassagne, 69424 Lyon Cedex 03, France (e-mail: antoineduclos@yahoo.fr)

Background: A control chart can help to interpret and reduce sources of variability in patient safety by continuously monitoring indicators. The aim of this study was to monitor the outcome of thyroid surgery using control charts.

Methods: Patients who had thyroid surgery during 2006–2007 were included in the study. Safety was monitored based on postoperative complications of recurrent laryngeal nerve palsy and hypocalcaemia. Indicators were extracted prospectively from the hospital information system and plotted each month on a P-control chart. Performance of the surgical team was also measured retrospectively for 2004–2005 (baseline period) to compare surgical outcomes before and after control chart implementation. Electromyographic monitoring of recurrent laryngeal nerves was not used, nor was calcium or vitamin D given routinely.

Results: The outcomes of 1114 thyroid procedures were monitored. Although the proportion of patients with recurrent laryngeal nerve palsy was similar for baseline and monitored periods (6.4 and 7.2 per cent respectively), there was a 35.3 per cent decrease in hypocalcaemia after implementation of control charts ($P < 0.001$). Complications almost doubled during a period when one surgeon was away and operating room renovations took place.

Conclusion: Outcome monitoring in thyroid surgery using control charts is useful for identifying potential issues in patient safety.

Presented to the 25th International Conference of the International Society for Quality in Health Care, Copenhagen, Denmark, October 2008 and to the 24th PCSI Working Conference, Lisbon, Portugal, October 2008

Paper accepted 5 January 2009

Published online in Wiley InterScience (www.bjs.co.uk). DOI: 10.1002/bjs.6418

Introduction

The outcome of thyroid surgery is usually assessed on the occurrence of two major adverse events: recurrent laryngeal nerve paralysis and hypoparathyroidism. The proportion of patients suffering from these complications at a single institution may vary because of changes in care, or for other reasons such as changes in data collection methods, case mix evolution or random factors¹. Health service performance monitoring involves regular assessment of indicators related to quality of care². The statistical process control is a decision support method that combines time series analysis with graphical presentation of data, called a Shewhart control chart. It was developed in the manufacturing industry³ and is increasingly applied in

research and healthcare improvement^{4,5}. The purpose of the present study was to monitor thyroid surgery outcomes using control charts.

Methods

All patients who had thyroid surgery at the authors' institution in the monitoring period (1 January 2006 to 31 December 2007) were considered for inclusion. Patient safety indicators were extracted prospectively from the hospital information system and plotted each month on a control chart during this period. To compare surgical outcomes before and after control chart implementation, the performance of the surgical team was determined

by a retrospective medical record review of 346 patients randomly selected from an earlier period (2004–2005: the baseline period).

Outcome indicators

Two indicators were used to monitor the safety of thyroid surgery. The first was the in-hospital proportion of immediate recurrent laryngeal nerve palsy in patients who had a unilateral or bilateral thyroid procedure. Postoperative vocal fold mobility was systematically assessed by anaesthetists using flexible transnasal laryngoscopy. When this was inconclusive or the voice was abnormal, laryngoscopy was repeated by an ear, nose and throat specialist. Electromyographic monitoring of recurrent laryngeal nerve was not used during the study period. The second indicator was the in-hospital proportion of immediate hypocalcaemia in patients who had bilateral thyroidectomy. Postoperative hypocalcaemia was defined as a total serum calcium level lower than 2 mmol/l in at least one measurement (reference range 2.25–2.60 mmol/l). No patients received calcium or vitamin D supplements before surgery, but hypocalcaemic patients did receive one or both of these after surgery as needed.

Control chart characteristics

Statistical process control is based on the use of a control chart to help the user distinguish between variation due to special causes (not inherent) and common causes (inherent). The control chart includes three horizontal lines (the central line, the upper control limit and the lower control limit) and plots measurements chronologically⁶. Longitudinal P-control charts were constructed in which each data point expressed the observed proportion of complications per month for each sample (samples could vary in size). The value for the central line was the mean proportion of complications in the monitoring period. Control and warning limits were calculated for each month based on two and three binomial-based standard deviations of the mean proportion of complications respectively. A special cause variation was defined as a single point outside the upper or lower control limits or as two of three successive points between a warning limit and a control limit on the same side of the central line⁵.

Statistical analysis

Outcomes of thyroid surgery were compared between the baseline (2004–2005) and monitoring (2006–2007) periods using the χ^2 test. Analyses were conducted using SAS[®]

Table 1 Patient characteristics

	Baseline period 2004–2005 (n = 346)	Monitoring period 2006–2007 (n = 1114)
Age (years)*	50.1 (6.0–78.4)	51.7 (8.8–93.5)
Women	277 (80.1)	877 (78.7)
Thyroid disease		
Malignant neoplasm	81 (23.4)	262 (23.5)
Non-toxic solitary nodule	45 (13.0)	164 (14.7)
Non-toxic multinodular goitre	160 (46.2)	563 (50.5)
Thyrotoxicosis with diffuse goitre	26 (7.5)	81 (7.3)
Other	34 (9.8)	44 (3.9)
Type of surgery		
Lobectomy	43 (12.4)	163 (14.6)
Total thyroidectomy	258 (74.6)	809 (72.6)
Thyroidectomy with lymph node resection	13 (3.8)	74 (6.6)
Completion thyroidectomy	32 (9.2)	59 (5.3)
Other	0 (0)	9 (0.8)

Values in parentheses are percentages unless indicated otherwise; *values are median (range).

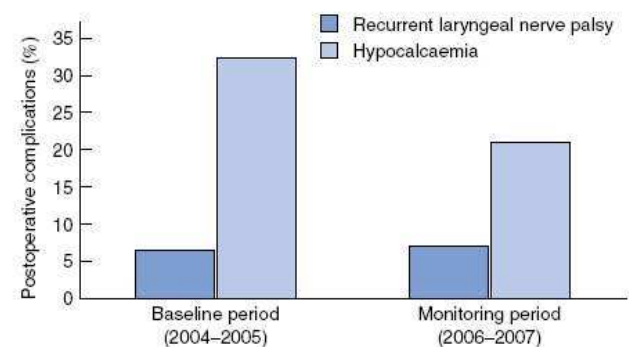


Fig. 1 Proportion of patients with recurrent laryngeal nerve palsy and hypocalcaemia (serum calcium level lower than 2 mmol/l) before and after control chart implementation

9.13 (SAS Institute, Cary, North Carolina, USA), and control charts were generated using Microsoft Office Excel[®] 2007 (Microsoft, Redmond, Washington, USA).

Results

The surgical team performed 1315 thyroid procedures between January 2006 and December 2007. Of these, only 1114 (84.7 per cent) were well documented and included in the study. Of the procedures, 942 (84.6 per cent) were bilateral thyroidectomies, and 262 (23.5 per cent) involved patients with thyroid carcinoma (Table 1). The proportion of recurrent laryngeal nerve palsy was similar in the two periods, but the proportion of hypocalcaemia

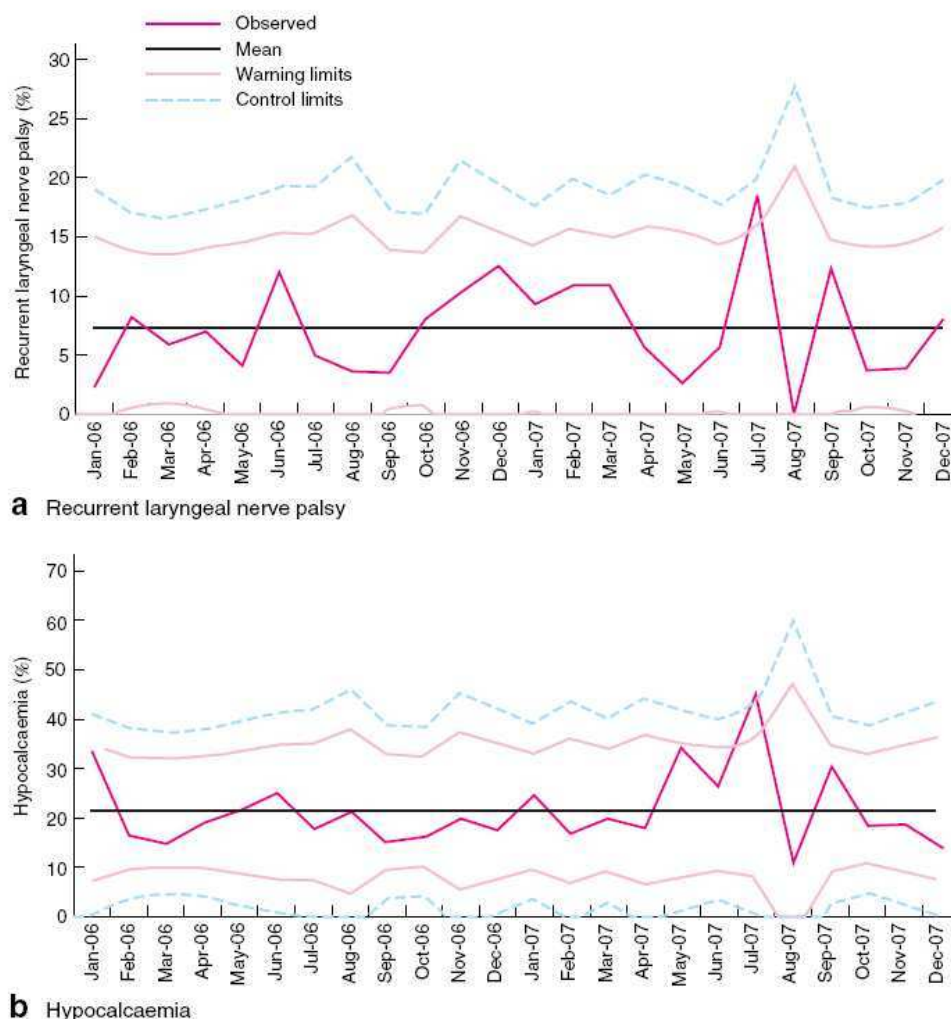


Fig. 2 Control charts for monthly proportions of postoperative complications: **a** recurrent laryngeal nerve palsy and **b** hypocalcaemia (serum calcium level lower than 2 mmol/l)

in the monitoring period fell by 35.3 per cent from 32.3 to 20.9 per cent ($P < 0.001$) (Fig. 1).

Control chart interpretation

Longitudinal P-control charts showed monthly proportions of postoperative complications (Fig. 2). Outcomes worsened during the third quarter of 2007: 13 patients had recurrent laryngeal nerve palsy and 28 had hypocalcaemia compared with seven and 17 events, respectively, as expected based on the mean proportions of complications from the whole 2-year period (without considering data from the third quarter of 2007).

An analysis of the special cause variation revealed two causes: first, an operating room was moved because of renovations (July to September 2007), leading to a

reduction in operating time available (from 42 to 32 h per week); second, one surgeon was absent in July 2007, although the number of patients having thyroid surgery did not decrease, so in July almost all procedures were performed by the only other surgeon available.

Discussion

After implementation of control charts used to monitor safety of thyroid surgery, this study found a significant reduction in postoperative hypocalcaemia. However, the observed proportions of recurrent laryngeal nerve palsy before and after implementation were similar.

The reduction in postoperative hypocalcaemia may represent a Hawthorne effect³ (increased caution under observation) or it may have resulted from indicator feedback⁷.

The heterogeneity in data collection between the baseline and monitoring periods may have also introduced variability in outcome measurements. Nevertheless, the modalities for detecting and coding postoperative complications remained unchanged, as the similarity in the proportions of recurrent laryngeal nerve palsy before and after chart implementation would support.

Observed improvements in preventing postoperative hypocalcaemia, as opposed to recurrent nerve palsy, probably reflect different degrees of surgeon vigilance with regard to the preservation of the parathyroid glands compared with the recurrent laryngeal nerve. Systematic dissection of the recurrent laryngeal nerve, a technique adopted by surgeons long ago, can be performed easily because of an almost uniform anatomy across patients. Conversely, there is great variability in the anatomy of the inferior parathyroid glands and their vascularization, so these glands require very meticulous dissection. It is likely that, because of the implementation of control charts, increased vigilance of surgeons mainly affected the preservation of the parathyroid glands.

Furthermore, the poor performance in the third quarter of 2007 appeared to be related to changes in the operating room and a surgeon's vacation, which placed the surgical team under great pressure. The physical environment and excessive workload are well known error-producing conditions⁸. In the operating theatre, fatigue combined with stress makes complex monitoring tasks particularly vulnerable to adverse events. The overactivity of a single surgeon was probably detrimental to the quality of thyroid surgery.

Although patient case mix and surgeon's skills constitute major determinants of successful surgery, the working environment may also impact surgical outcomes. The primary concern should be to guarantee that surgeons can operate under satisfactory conditions and that they are not given excessive workloads. Specific measures such as increasing equipment availability and use (such as recurrent laryngeal nerve monitoring by neurostimulation⁹), team coordination and streamlined rostering may improve outcomes. In conjunction with interventions to improve care, efforts should be made to conduct more reliable surveillance of surgical performance. Performance monitoring using outcome indicators requires that the observed variations reflect true variations in surgical safety. When validating control charts of crude proportions of complications, it is important to ensure that this tool detects variations equally to control charts based on case mix-adjusted proportions of complications. Real-time monitoring of individual surgeon performance with cumulative sum charts (CUSUM)¹⁰

would help to detect a small shift in surgical practice before too many complications occurred¹¹.

This study suggests that monitoring outcomes in thyroid surgery using control charts is feasible and useful in the daily practice of a surgical ward. Shared experiences among surgical teams who have implemented this tool can improve patient safety.

Acknowledgements

The authors are grateful to Philippe Messy for his helpful contribution to the study, and to Anne-Marie Scott and Stephen Martin for critical revision of the manuscript. The authors declare no conflict of interest.

References

- 1 Mant J. Process *versus* outcome indicators in the assessment of quality of health care. *Int J Qual Health Care* 2001; **13**: 475–480.
- 2 Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* 2004; **363**: 1147–1154.
- 3 Best M, Neuhauser D, Walter A, Shewhart, 1924, and the Hawthorne factory. *Qual Saf Health Care* 2006; **15**: 142–143.
- 4 Thor J, Lundberg J, Ask J, Olsson J, Carli C, Härenstam KP *et al*. Application of statistical process control in healthcare improvement: systematic review. *Qual Saf Health Care* 2007; **16**: 387–399.
- 5 Benneyan JC, Lloyd RC, Plsek PE. Statistical process control as a tool for research and healthcare improvement. *Qual Saf Health Care* 2003; **12**: 458–464.
- 6 Marshall T, Mohammed MA. Differences in clinical performance. *Br J Surg* 2002; **89**: 948–949.
- 7 Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2006; (2)CD000259.
- 8 Dean B, Schachter M, Vincent C, Barber N. Causes of prescribing errors in hospital inpatients: a prospective study. *Lancet* 2002; **359**: 1373–1378.
- 9 Dralle H, Sekulla C, Haerting J, Timmermann W, Neumann HJ, Kruse E *et al*. Risk factors of paralysis and functional outcome after recurrent laryngeal nerve monitoring in thyroid surgery. *Surgery* 2004; **136**: 1310–1322.
- 10 Woodall WH. The use of control charts in health-care and public-health surveillance. *J Qual Technol* 2006; **38**: 89–104.
- 11 Lim TO, Soraya A, Ding LM, Morad Z. Assessing doctors' competence: application of CUSUM technique in monitoring doctors' performance. *Int J Qual Health Care* 2002; **14**: 251–258.

5.2. Travail méthodologique sur le développement d'une carte adaptée au suivi des complications de la chirurgie thyroïdienne (*articles 2 & 3*)

La carte de contrôle est un outil initialement développé dans l'industrie manufacturière afin d'y être appliquée sur des chaînes de production à très haut volume et pour lesquelles un nombre limité de facteurs de variabilité est susceptible d'influencer la qualité du produit final. Son utilisation dans le domaine de la santé peut s'avérer délicate, compte tenu des faibles effectifs de patients et d'évènements dont on dispose parfois pour évaluer une prise en charge ainsi que des nombreux facteurs qui peuvent concourir à son résultat. Bien que la carte de Shewhart soit relativement facile d'utilisation, sa construction implique certaines précautions et son interprétation doit être prudente afin d'éviter des conclusions hâtives pouvant conduire à des décisions inappropriées. En particulier, la manière dont ses limites sont calculées peut être déterminante sur sa performance à détecter des variations anormales de l'indicateur suivi.

Cette seconde partie du projet, davantage méthodologique, portait sur l'outil en lui-même et sur les modalités de calcul de ses limites. Nous avons souhaité développer une carte de Shewhart adaptée au suivi mensuel des complications de la chirurgie thyroïdienne. Deux travaux ont été menés conjointement pour y parvenir.

Considérant que le nombre de patients opérés comme le nombre de complications observées chaque mois pouvait s'avérer incompatible avec l'application d'une loi de distribution normale, nous avons d'abord comparé la détection de causes spéciales entre une carte à limites approximées (reposant sur une distribution normale) et une carte à limites exactes (reposant sur une distribution binomiale). Utilisant des données fictives, cet exemple suggérait que le mode de calcul des limites était déterminant sur la capacité de l'outil à détecter une cause spéciale de variation. Si les conditions pour une approximation normale de la distribution binomiale n'étaient pas réunies, il était fortement recommandé d'interpréter les variations d'indicateur sur la carte à l'aide de limites exactes.

Nous avons ensuite jugé qu'avant de considérer les variations de complications mensuelles observées en chirurgie thyroïdienne comme étant le reflet de la qualité des pratiques chirurgicales, il convenait de s'assurer qu'elles n'étaient pas liées aux caractéristiques des patients opérés. Une validation des cartes de contrôle à limites brutes a donc été menée en comparant la détection des variations d'indicateurs obtenue sur ces cartes à celle obtenue sur des cartes à limites ajustées. Portant sur le résultat de 1 405 interventions

réalisées de janvier 2006 à mai 2008, cette comparaison a montré une concordance remarquable entre cartes à limites brutes et ajustées. Néanmoins, la seule cause spéciale de variation survenue n'était détectée que par la carte à limites ajustées, suggérant une utilisation pertinente de cette carte en complément de celle à limites brutes lorsque l'indicateur se trouvait à proximité des limites de contrôle.

5.2.1. Performance d'une carte à limites exactes vs. approximées

Antoine Duclos, Nicolas Voirin.

The p-control chart: a tool for care improvement.

International Journal for Quality in Health Care 2010 (sous presse).

Le second article en cours de révision et reproduit ci-après constitue un exposé didactique visant à fournir au chirurgien comme au chercheur ou à l'étudiant des bases méthodologiques sur la manière de développer et interpréter une carte de contrôle en pratique courante, ainsi que d'implémenter cet outil avec succès sur le terrain et d'en restituer les résultats. En particulier, il présente différentes méthodes de calcul des limites de contrôle, insistant sur la nécessité de privilégier la construction de cartes à limites exactes.

The p-control chart: a tool for care improvement

Antoine Duclos¹ and Nicolas Voirin²

¹ *Hospices Civils de Lyon, Pôle Information Médicale Evaluation Recherche; Université de Lyon, EA Santé-Individu-Société 4129, Lyon, France.*

² *Université de Lyon 1, Laboratoire de Biométrie et Biologie Evolutive; Hospices Civils de Lyon, Hôpital Edouard Herriot, Service d'Hygiène, Epidémiologie et Prévention, Lyon, France.*

Abstract

Background: The p-chart is a user-friendly tool for monitoring adverse events. By converting data into knowledge, it is helpful in interpreting and reducing sources of variability in care. Certain basics for developing expertise to use p-charts correctly are necessary.

Purpose: This paper provides key elements on how to develop and interpret a p-chart for clinical practice, how to successfully integrate this tool within a comprehensive approach, and how to report a study based on p-chart utilisation.

P-chart building: The p-chart combines time series analysis with a graphical presentation of data by plotting successive indicator measurements in chronological order. The pragmatic choice of well-defined indicators to be monitored is essential. Exact control limits based on the binomial distribution and the incorporation of risk adjustment represent important contributions for further improving the tool's performance in healthcare settings.

P-chart implementation: The solution needed to reduce adverse events is not available from measurement alone. The success of routine introduction of p-chart requires investigating the causes of indicator variations and trying quality improvement initiatives. It must be supported by strong management leadership within an atmosphere of constructive evaluation.

Perspectives: The implementation of the p-chart into clinical practice encourages practitioners to continuously undertake a critical examination of the care delivered. Nearly a century after it was created in the manufacturing industry, the control chart now contributes to improving the quality of healthcare processes and patient safety.

Introduction

The *p*-control chart is a graphical tool developed in industry to interpret and reduce sources of variability in manufacturing processes. It is now increasingly applied in healthcare for continuous quality control and quality improvement research. There is strong interest in implementing *p*-charts in clinical practice to monitor adverse events and guide initiatives for the improvement of patient safety. Although the *p*-chart is easy to use, its design requires certain precautions and it must be interpreted carefully in order to avoid erroneous conclusions. This paper provides key elements to physicians, nurses, managers, students or researchers of quality on how to develop and interpret a *p*-chart for clinical practice, how to successfully integrate this tool within a comprehensive control chart program, and how to report a study based on *p*-chart utilisation. In order to help the non-experts in

designing the *p*-control chart correctly, additional materials are appended, including a statistical appendix and an online spreadsheet with formulas for plotting *p*-chart.

Basic principles of the control chart

The statistical control chart concept was first developed in the 1920s by Dr Walter Shewhart in order to improve the reliability of telephone transmission systems.¹ The concept resulted from the observation that operators often overreact and make inappropriate changes in settings in response to indicator variations that are simply random.² Such decisions are wasteful and introduce more variation in the process, making the system less reliable. Shewhart's theory of variation states that quality is inversely proportional to variability and that understanding the variability of some indicators could teach the operator about when and how to reduce it.

The implementation of control charts in health services to monitor adverse events was advocated by Edwards Deming in 1942.³ It has since been used in a wide range of settings and specialties, which suggests a broad applicability in the healthcare context.⁴ Based on the monitoring of patients' health outcome indicators, the control chart detects significant changes over time in patient safety. In the field of quality control, the routine use of control charts is helpful in interpreting and reducing sources of variability in care.⁵ By converting data into knowledge, it guides multidisciplinary teams towards the most appropriate action(s) for continuous improvement.⁶ In health services research, the control chart represents a low-cost and robust method for rapidly assessing change initiatives in care.⁷ It can be viewed as a quasi-experimental study design, which is complementary to randomised controlled trials in providing evidence on the impact of safety improvement interventions.⁸ Control charts have also been successfully used in other areas such as public health surveillance in communities⁹ or in hospitals,¹⁰ benchmarking of hospital performance,^{6,11} or monitoring of clinical variables at an individual patient level.¹²

Though relatively simple, the control chart is a powerful tool for decision support. It combines time series analysis with a graphical presentation of data by plotting successive indicator measurements in chronological order. By distinguishing special causes from common causes of variation, the control chart categorises variation according to the action needed to reduce it.¹³ Special causes are supposed to reflect substantial variation in care that deserves further investigation, whereas variation related to common causes is expected to arise due to other misleading factors, including random events. Three horizontal lines are drawn on the chart to determine whether or not care is in statistical control; they are termed the central line, the upper control limit (UCL), and the lower control limit (LCL). An indicator data point lying outside of the control limits suggests that some special cause of variation has been detected and that care is out of control. This requires finding and acting on one or more assignable causes to reduce variation. After the elimination of all special causes of variation, care should be subject only to common-cause

variation related to unknown or unmeasured factors. In this case, if the level of performance is still regarded as unsatisfactory, the classical way to further improve care is to reorganise the whole process. This implies choosing simplicity in the restructuring of care, considering that simple systems would be more reliable than complex ones.²

How to build a *p*-control chart

The *p*-chart, where *p* stands for proportion, is useful for the routine monitoring of a binary outcome, such as the occurrence of an adverse event (e.g., postoperative complications). Although the *p*-chart was conceived to be user-friendly for non-experts, previous knowledge is required for design.

In creating a *p*-chart, both the sample size and the frequency of sampling must first be specified before plotting the observed proportion of adverse events in each successive sample.¹⁴ A choice should be made between large samples at wider intervals (aiming to detect small shifts in the monitoring of rare adverse events) and small samples at shorter intervals (aiming to detect shifts as soon as possible for real-time monitoring). In practice, *p*-charts are usually displayed with at least 20 consecutive samples that can be of variable size, as it is relatively common to examine samples based on every patient cared for over some convenient period of time (e.g., the number of patients operated on each month).

Second, a central line is drawn, which corresponds to the overall proportion of adverse events across all samples, \bar{p} (e.g., the mean proportion of postoperative complications).

Third, the control limits are traditionally positioned at a distance of three standard deviations (SD) around the central line and the detection of special-cause variation depends on finding a single point outside of the control limits. This decision rule ensures an optimal balance between the tool's sensitivity in detecting signals (i.e., avoid mistaking a special cause for a common cause) and its specificity in avoiding false alarms (i.e., avoid mistaking a common cause for a special cause). Accordingly, 99.73% of all points are expected to fall within 3 SD from the mean if the process is stable (common-cause variation), with the remaining 0.27% falling more than 3

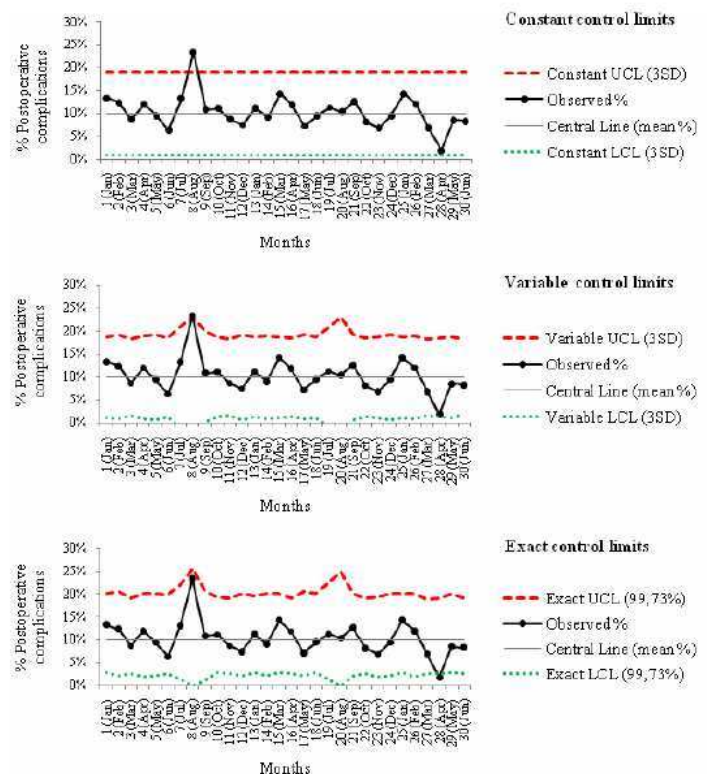
SD away from the mean (special-cause variation).¹⁵

There are three main approaches in setting control limits on a p -chart with a variable sample size (see Appendix for detailed formulas). The first one consists in calculating constant control limits based on the average of the sample sizes. This assumes that the sizes of successive samples do not vary greatly. However, in the case of an unusually large variation in the size of a particular sample or if an indicator is positioned close to the approximate limits, then interpretation must be conducted cautiously.¹⁶ To avoid such pitfalls, there is a second approach, which consists in determining variable control limits for each sample based on its specific size, n_i . In this case, the control limits will be drawn in stair-steps to reflect the changes in sample size over time. The more the sample size increases, the closer to the central line the limits will be. Generally, calculations of the p -chart limits are based on a normal approximation of the binomial distribution. This approximation is acceptable as long as $n_i \bar{p}(1-\bar{p}) > 5$ and $0.1 \leq \bar{p} \leq 0.9$. If these conditions are not satisfied, a third approach based on the calculation of exact control limits should be adopted, directly using the binomial distribution.¹³

The modalities for setting the control limits may have a great influence on the tool's performance, as shown in Figure 1. The p -charts displaying the three types of control limits described above were designed using Microsoft Office Excel®, so that the reader can replicate the method by accessing the on-line supplementary material containing fictive data, parameter calculations, and chart plotting. Many other popular spreadsheets or statistical software can also be used to build control charts easily.¹⁷ Each data point expresses the observed proportion of postoperative complications per month for 30 successive samples (Table 1). The central line is positioned at 10% and control limits are set at 3 SD from the central line. The conditions for a normal approximation were not satisfied for several points (see on-line data), making the control chart based on exact limits more reliable. When comparing the three charts, two points should be discussed specifically. No

special-cause variation was detected at month 8 by the p -chart based on exact limits, whereas it was erroneously detected by the p -charts based on constant and variable approximate limits. The single special-cause variation needing to be investigated was detected at month 28 only by the p -chart based on exact limits, but not by the two other charts.

Figure 1. Three p -charts with different setting of lower (LCL) and upper (UCL) control limits



In the end, if sample sizes vary, one should always prefer interpretation with variable limits rather than constant limits. Furthermore, if the conditions for normal approximation of the binomial distribution are not satisfied, it is recommended to base interpretation on the exact limits. Nevertheless, the key for success in using control chart should be to strike a balance between statistical correctness and special attention to every changes occurring in the delivery of care. The pragmatic use of control chart implies to interpret rigorously the indicator variations as well as to know what happens in the practical field, before and after measurement.

Table 1. Dataset for plotting *p*-charts

Month	No. of complications	No. of surgical procedures
1	14	105
2	12	97
3	10	115
4	12	100
5	9	95
6	7	111
7	9	68
8	11	47
9	9	83
10	12	108
11	10	115
12	7	94
13	12	107
14	9	99
15	15	105
16	13	110
17	7	97
18	10	105
19	8	71
20	5	48
21	12	95
22	9	110
23	7	103
24	9	95
25	15	105
26	12	100
27	8	116
28	2	110
29	9	105
30	10	120
Total	294	2939

How to develop a *p*-chart program

Box 1 provides the essential points that should be considered in a protocol aiming at implementing a *p*-chart program. To be successful, the use of *p*-charts must be supported by strong management leadership and by mechanisms for communicating results throughout the organisation. It should be integrated within a multidisciplinary team approach focusing on a strategic process of care.¹⁴ High-frequency and standardised care in relatively homogeneous patient groups should be primarily selected for *p*-chart control. The organisation should have a high potential for quality improvement and changes in care should be proposed in order to be tested. Another prerequisite implies the pragmatic choice of well-defined indicators to be monitored, with the ultimate goal of reducing their variability over time and achieving better outcomes.

Box 1. Implementing a *p*-chart program

- 1) Appointment of a motivated project head within the team to be placed under observation
- 2) Choice of care process to be put under control (high volume procedure with critical outcome and great potential for improvement)
- 3) Setting quality improvement objectives to be accomplished (reduction of outcome variability, performance level to achieve or maintain)
- 4) Selection of indicator to be monitored (based on its clinical relevance, construction feasibility and validity) and its calculation formula (ratio with monitored event as the numerator and exposed population as the denominator)
- 5) Selection of computerised data collection (standardised, continuous and exhaustive cases registration) and extraction systems (automatic data processing based on user-friendly software)
- 6) Setting control chart parameters (sample sizes and sampling frequency, central line position and width of control limits) and explaining interpretation rules
- 7) Selection of modalities for identification of special causes (logbook, chart restitution meetings or others quality control tools)
- 8) Detailing specific actions that could be tested with the aim of improving patient safety
- 9) Planning the periodicity and modalities of chart restitution within the organisation
- 10) Editing the timetable and assigning tasks among the different actors involved

Appropriate logistic support is necessary for developing a *p*-chart program, which may quickly yield cost savings by avoiding adverse events. Ideally, *p*-chart development requires a computerised system for data management and it must be implemented within a sustainable approach based on the Plan-Do-Check-Act cycle.¹⁴ Indeed, the solution needed to reduce adverse events is not available from measurement alone and requires local action to systematically explore the core reasons for the specific pattern of outcomes.¹⁸ If special causes of variation are detected by the *p*-chart, they must be investigated by “detective work” on the process. This involves appropriate methods or quality control tools,¹⁷ such as using a logbook in which all changes in care are continually reported. The holding of periodic meetings, during which control charts are interpreted by healthcare workers, may also be useful to generate knowledge about the root causes of observed indicators variations. Once special causes have been identified, some suitable actions must be conducted with the aim of conserving or eliminating these unusual

sources of variability according to the observed improvement or impairment, respectively, in care safety. Then, after obtaining a stable process, concrete interventions aiming at care improvement must be successively implemented and tested. Furthermore, to implement a *p*-chart program most effectively, it is essential to have an atmosphere of constructive evaluation rather than to judge the individual performance of the professionals whose outcomes are being assessed.¹⁹

How to report a study based on *p*-chart utilisation

When reporting a study based on the use of a *p*-chart, authors must clearly state certain information to help the reader in judging the rigor of the employed methodology (Box 2). Special attention should be paid in describing the modalities of identification of and actions taken against special causes of variation, as well as in detailing quality improvement interventions that have been assessed. Benefits and limitations should also be reported, along with barriers and facilitating factors related to the implementation of a *p*-chart program.

Box 2. Check-list for reporting a *p*-chart program^{4,12,20}

- 1) Study objective (tutorial, quality control, assessing the impact of interventions on care, individual patient monitoring, public health surveillance or performance monitoring)
- 2) Study design (retrospective or prospective, observational or quasi-experimental controlled before-and-after study)
- 3) Study setting (country, healthcare sector, medical or surgical specialty) and units under observation (hospital ward, primary care centre, surgical team, single clinician or patient)
- 4) Process of care under control
- 5) Monitoring period and number of samples examined
- 6) Modalities of construction of monitored indicators
- 7) Modalities of data collection and extraction systems (data sources, type of software used, data quality and exhaustiveness)
- 8) Control chart parameters (sample sizes and sampling frequency, central line and limit settings, adjustment for case-mix or other confounding factors) and rules of interpretation
- 9) Complementary methods and quality control tools for identifying special causes
- 10) Specific safety improvement actions that have been tested

Conclusions

Monitoring a healthcare process is different from monitoring a manufacturing process. The rarity of events and confounding factors such as patient case-mix represent fundamental differences. Considering that the probability of adverse events may vary considerably across patients undergoing particular care interventions, risk adjustment or stratification is useful to enable correct analysis of data from heterogeneous populations. Accordingly, the use of adjusted *p*-charts seems to be a valuable approach to improve the tool's reliability in healthcare settings.^{21,22} Warning limits (2 SD around the central line) can also be added to improve the sensitivity of the control chart,¹⁵ but this tends to increase the risk of false alarms in detecting special-cause variation.¹³ Furthermore, the cumulative sum chart (CUSUM) may be helpful in overcoming the limitations of *p*-chart sensitivity in clinical practice, as it performs fairly well in detecting small changes when monitoring rare adverse events.^{11,19}

The implementation of the *p*-chart into clinical practice encourages practitioners to continuously undertake a critical examination of the care delivered. Deming predicted in the 1980s that *"another half-century may pass before the full spectrum of Dr. Shewhart's contributions has been revealed in liberal education, science, and industry"*.²³ Nearly a century after it was created in the manufacturing industry, the control chart now contributes to improving the quality of healthcare processes and patient safety.

Acknowledgements

None.

Funding Support

None declared.

Conflicts of Interest

The authors have no potential conflicts of interest to be disclosed.

References

1. Shewhart WA. *Economic control of quality of manufactured product*. New York: Van Nostrand, 1931.
2. Berwick DM. Controlling variation in health care: a consultation from Walter Shewhart. *Med Care* 1991; 29(12): 1212-25.
3. Deming WE. On a classification of the problems of statistical inference. *J Am Stat Ass* 1942; 37(218): 173-85.
4. Thor J, Lundberg J, Ask J et al. Application of statistical process control in healthcare improvement: systematic review. *Qual Saf Health Care* 2007; 16(5): 387-99.
5. Duclos A, Touzet S, Soardo P et al. Quality monitoring in thyroid surgery using the Shewhart control chart. *Brit J Surg* 2009; 96(2): 171-4.
6. Mohammed MA, Cheng KK, Rouse A et al. Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons. *Lancet* 2001; 357(9254): 463-7.
7. Norberg A, Christopher NC, Ramundo ML et al. Contamination rates of blood cultures obtained by dedicated phlebotomy vs intravenous catheter. *JAMA* 2003; 289(6): 726-9.
8. Diaz M, Neuhauser D. Pasteur and parachutes: when statistical process control is better than a randomized controlled trial. *Qual Saf Health Care* 2005; 14(2): 140-3.
9. Hanslik T, Boelle PY, Flahault A. The control chart: an epidemiological tool for public health monitoring. *Public Health* 2001; 115(4): 277-81.
10. Morton AP, Whitby M, McLaws ML et al. The application of statistical process control charts to the detection and monitoring of hospital-acquired infections. *J Qual Clin Pract* 2001; 21(4): 112-7.
11. Woodall WH. The use of control charts in health-care and public-health surveillance, *J Qual Technol* 2006; 38(2): 89-104.
12. Tennant R, Mohammed MA, Coleman JJ et al. Monitoring patients using control charts: a systematic review. *Int J Qual Health Care* 2007; 19(4): 187-94.
13. Mohammed MA, Worthington P, Woodall WH. Plotting basic control charts: tutorial notes for healthcare practitioners. *Qual Saf Health Care* 2008; 17(2): 137-45.
14. Montgomery, Douglas C. *Statistical Quality Control: A Modern Introduction, International Student Version, 6th ed*. Hoboken, NJ: Wiley, 2008.
15. Benneyan JC, Lloyd RC, Plsek PE. Statistical process control as a tool for research and healthcare improvement. *Qual Saf Health Care* 2003; 12(6): 458-64.
16. Ryan, Thomas P. *Statistical methods for quality improvement, 2nd ed*. New York: Wiley, 2000.
17. Lighter D. *Advanced Performance Improvement in Health Care: Principles and Methods*. Sudbury, MA: Jones and Bartlett Publishers, 2010.
18. Werner RM, McNutt R. A new strategy to improve quality: rewarding actions rather than measures. *JAMA* 2009; 301(13): 1375-7.
19. Rogers CA, Reeves BC, Caputo M, et al. Control chart methods for monitoring cardiac surgical performance and their interpretation. *J Thorac Cardiovasc Surg* 2004; 128(6): 811-9.
20. Biau DJ, Resche-Rigon M, Godiris-Petit G et al. Quality control of surgical and interventional procedures: a review of the CUSUM. *Qual Saf Health Care* 2007; 16(3): 203-7.
21. Duclos A, Voirin N, Touzet S, et al. Crude versus case-mix-adjusted control charts for safety monitoring in thyroid surgery. *Qual Saf Health Care* 2009; IN PRESS.
22. Amin SG. Control charts 101: a guide to health care applications. *Qual Manag Health Car* 2001; 9(3): 1-27.
23. Shewhart WA. *Statistical Method from the Viewpoint of Quality Control*. Washington DC: Graduate School of the Department of Agriculture, 1939.

Appendix

Detailed formulas for building a p -control chart are given. An example for setting constant, variable and exact control limits is provided, based on the sample of surgical procedures referring to the tenth month in Table 1 and Figure 1. The corresponding MS Excel formulas can be read and used in the appended spreadsheet on-line.

Calculation of the observed proportion of adverse events for plotting points

The observed proportion of adverse events in each sample is $p_i = \frac{x_i}{n_i}$, where n_i is the sample size and x_i is the number of adverse events in the sample i .

- Example for the calculation of the observed proportion of postoperative complications with the 10th sample: $p_{10} = \frac{12}{108} = 11.1\%$.

Calculation of the observed overall proportion of adverse events for plotting the central line

The mean proportion of adverse events is $\bar{p} = \frac{\sum_{i=1}^k x_i}{\sum_{i=1}^k n_i}$, where k is the number of samples.

- Example for the calculation of the mean observed proportion of postoperative complications: $\bar{p} = \frac{14+12+\dots+9+10}{105+97+\dots+105+120} = \frac{294}{2,939} = 10.0\%$.

Calculation of the lower control limit (LCL) and the upper control limit (UCL)

- (i) The constant control limits are calculated using $\bar{p} \pm 3\sqrt{\frac{\bar{p} \times (1 - \bar{p})}{\bar{n}}}$, where \bar{n} is the average sample size

calculated using $\bar{n} = \frac{1}{k} \sum_{i=1}^k n_i$.

- Example for the calculation of the average sample size: $\bar{n} = \frac{1}{30} \times (105 + 97 + \dots + 105 + 120) = 98$.

- Example for the calculation of the constant LCL:

$$\text{Constant LCL} = 0.1 - 3\sqrt{\frac{0.1 \times (1 - 0.1)}{98}} = 0.1 - 3 \times 0.03 = 0.9\%.$$

- Example for the calculation of the constant UCL:

$$\text{Constant UCL} = 0.1 + 3\sqrt{\frac{0.1 \times (1 - 0.1)}{98}} = 0.1 + 3 \times 0.03 = 19.1\%.$$

- (ii) The variable control limits are calculated using $\bar{p} \pm 3\sqrt{\frac{\bar{p} \times (1 - \bar{p})}{n_i}}$, where n_i is the size of each sample.

- Example for the calculation of the variable LCL with the 10th sample:

$$\text{Variable LCL}_{10} = 0.1 - 3\sqrt{\frac{0.1 \times (1 - 0.1)}{108}} = 0.1 - 3 \times 0.029 = 1.34\%.$$

- Example for the calculation of the variable UCL with the 10th sample:

$$\text{Variable UCL}_{10} = 0.1 + 3\sqrt{\frac{0.1 \times (1 - 0.1)}{108}} = 0.1 + 3 \times 0.029 = 18.7\%.$$

- (iii) The exact control limits can be calculated using the cumulative binomial distribution expressed statistically

by $F(x_i; \bar{p}, n_i) = P(X_i \leq x_i) = \sum_{k=0}^{x_i} \binom{n_i}{k} \bar{p}^k (1 - \bar{p})^{n_i - k}$. For a 3 SD control limit strategy, the exact LCL is

the value of x_i that satisfies $P(X_i \leq x_i) = \frac{1 - 0.9973}{2} = 0.00135$ and the exact UCL is the value of x_i that

satisfies $P(X_i \leq x_i) = \frac{1 + 0.9973}{2} = 0.99865$. In practice, this can be done using the CRITBINOM function of MS Excel.

- Example for the calculation in MS Excel of the exact LCL with the 10th sample:

$$\text{Exact LCL}_{10} = \frac{\text{CRITBINOM}(n_i; \bar{p}; 0.00135)}{n_i} = \frac{\text{CRITBINOM}(108; 0.1; 0.00135)}{108} = 2.8\%.$$

- Example for the calculation in MS Excel of the exact UCL with the 10th sample:

$$\text{Exact UCL}_{10} = \frac{\text{CRITBINOM}(n_i; \bar{p}; 0.99865)}{n_i} = \frac{\text{CRITBINOM}(108; 0.1; 0.99865)}{108} = 19.4\%.$$

5.2.2. Performance d'une carte à limites ajustées vs. brutes

Antoine Duclos, Nicolas Voirin, Sandrine Touzet, Pietro Soardo, Anne-Marie Schott, Cyrille Colin, Jean-Louis Peix, Jean-Christophe Lifante.

Crude versus case-mix-adjusted control charts for safety monitoring in thyroid surgery.

Quality and Safety in Health Care 2010 (sous presse).

Le troisième article en cours de publication et reproduit ci-après présente une comparaison de la détection des variations observées sur les indicateurs de complications de la chirurgie thyroïdienne, entre une carte de contrôle à limites brutes et une carte de contrôle à limites ajustées sur le *case-mix* des patients opérés. Il recommande l'usage complémentaire d'une carte à limites ajustées en cas de difficultés d'interprétation de la carte à limites brutes.

Crude versus case-mix-adjusted control charts for safety monitoring in thyroid surgery

Antoine Duclos,^{1,2} Nicolas Voirin,^{3,4} Sandrine Touzet,^{1,2} Pietro Soardo,⁵
Anne-Marie Schott,^{1,2} Cyrille Colin,^{1,2} Jean-Louis Peix,⁵ Jean-Christophe Lifante⁵

¹Hospices Civils de Lyon, Pôle Information Médicale Évaluation Recherche, Lyon, France

²Université de Lyon, EA Santé-Individu-Société 4129, Lyon, France ³Université de Lyon 1, Laboratoire de Biométrie et Biologie Évolutive, Lyon, France

⁴Hospices Civils de Lyon, Hôpital Édouard Herriot, Service d'Hygiène, Épidémiologie et Prévention, Lyon, France

⁵Hospices Civils de Lyon, Centre Hospitalier Lyon Sud, Service de Chirurgie Générale et Endocrinienne, Pierre Bénite, France

Correspondence to

Dr Antoine Duclos, Hospices Civils de Lyon, Pôle d'Information Médicale Évaluation Recherche, Unité d'Étude des Pratiques Professionnelles, 162 avenue Lacassagne, F-69424 Lyon Cedex 03, France; antoineduclos@yahoo.fr

Previous communication The study's main results were presented at the 24th Patient Classification Systems International Working Conference (Lisbon, Portugal, 8–11 October 2008).

Accepted 24 June 2009

ABSTRACT

Introduction Patient-safety monitoring based on health-outcome indicators can lead to misinterpretation of changes in case mix. This study aimed to compare the detection of indicator variations between crude and case-mix-adjusted control charts using data from thyroid surgeries.

Methods The study population included each patient who underwent thyroid surgery in a teaching hospital from January 2006 to May 2008. Patient safety was monitored according to two indicators, which are immediately recognisable postoperative complications: recurrent laryngeal nerve palsy and hypocalcaemia. Each indicator was plotted monthly on a p-control chart using exact limits. The weighted κ statistic was calculated to measure the agreement between crude and case-mix-adjusted control charts.

Results We evaluated the outcomes of 1405 thyroidectomies. The overall proportions of immediate recurrent laryngeal nerve palsy and hypocalcaemia were 7.4% and 20.5%, respectively. The proportion of agreement in the detection of indicator variations between the crude and case-mix-adjusted p-charts was 95% (95% CI 85% to 99%). The strength of the agreement was $\kappa=0.76$ (95% CI 0.54 to 0.98). The single special cause of variation that occurred was only detected by the case-mix-adjusted p-chart.

Conclusions There was good agreement in the detection of indicator variations between crude and case-mix-adjusted p-charts. The joint use of crude and adjusted charts seems to be a reasonable approach to increase the accuracy of interpretation of variations in outcome indicators.

of variation are detected. However, there is still debate about how to use control charts to monitor the outcomes of care, specifically regarding the necessity of adjusting for changes in patient case mix. Some argue that case-mix adjustment remains complex,⁴ may not be essential for longitudinal monitoring of outcomes⁵ or can lead to the erroneous conclusion of an unbiased measure.⁶ However, a control chart adjusted for case mix could theoretically improve the detection of special cause variations by controlling for biases that result from changes in case mix over time.^{7,8} In thyroid surgery, reported complication rates may vary depending on the extent of resection and the severity of the pathological thyroid condition, as well as the patient's characteristics.^{9–11} This study aimed to compare the detection of indicator variations between crude and case-mix-adjusted control charts in monitoring the safety of thyroid surgery.

METHODS

Population and data collection

A prospective survey was conducted in the department of endocrine surgery in a large teaching hospital.¹² The study population included every patient who underwent thyroid surgery during a 29-month period between 1 January 2006 and 31 May 2008. The surgical team under observation included three surgeons. Specific information about each thyroid surgery and its related complications was collected by the surgeon in charge of the patient using a standardised form. Other information was obtained from the hospital information system, which contains standard discharge abstracts, including compulsory data on the patients' demographics and the thyroid disease diagnoses.

Indicators and control chart characteristics

Two indicators were used to monitor the safety of thyroid surgery. The first indicator was the in-hospital proportion of immediate postoperative recurrent laryngeal nerve palsy, which was systematically assessed by flexible transnasal laryngoscopy among patients who underwent a lobectomy or a bilateral thyroidectomy. The second indicator was the in-hospital proportion of immediate postoperative hypocalcaemia, which was rigorously defined by a serum calcium level lower than 2 mmol/l only among patients who underwent a bilateral thyroidectomy.

Each indicator was extracted from the hospital information system and plotted on a Shewhart p-control chart.¹ Crude and case-mix-adjusted p-charts were constructed such that each data point

INTRODUCTION

The statistical control chart concept is a method of decision support combining time-series analysis with a graphical presentation of data.¹ It is based on the use of a control chart that guides the users towards appropriate action for improvement according to the nature of the variability.² The control chart is a well-documented tool to monitor and improve the safety of healthcare processes.³ It is helpful in interpreting and reducing sources of indicator variability by distinguishing special causes of variation from common causes. Special cause variations reflect substantial disparities in patient safety that deserve further investigation and action. In contrast, common cause variations arise due to misleading factors, including random events.

There is strong interest in integrating control charts into clinical practice, particularly to alert multidisciplinary clinical teams when special causes

expressed the observed proportion of complications per month for samples of variable size, as follows:

- The central line value of the crude p-chart was constant and was determined based on the overall proportion of complications. Exact limits were calculated for each month using the binomial-based SD of the overall proportion of complications.
- The central line value of the case-mix-adjusted p-chart varied monthly and was established based on the expected proportion of complications per month. Exact limits were calculated for each month using the binomial-based SD of the expected proportions of complications.

When a special cause variation was detected by the p-charts, after checking the completeness and quality of the data, a systematic investigation was undertaken to identify the cause using two complementary methods. We first looked at the logbook, in which all changes in care processes were continually reported by surgeons. Every 4 months, a multidisciplinary meeting was also conducted to discuss the observed variations in monitored indicators on control charts.

Statistics

Case-mix adjustment was based on the patients' gender and age, thyroid disease diagnosis and type of thyroidectomy.^{9–11} Case-mix-adjusted ORs for each postoperative complication were calculated using a multivariate logistic regression with all previous variables entered in the model. The expected proportion of each complication was then calculated for each month based on the model estimates. Final models were assessed for potential interactions. A finding of non-significance ($p > 0.05$ and the closest to 1) following the Hosmer–Lemeshow test was interpreted as an adequate fit of the model to the data.¹³ The discriminatory power of the models was measured using the c-statistics (Receiver Operating Characteristic curve score). Values less than 0.7 were considered to show poor discrimination.¹⁴

The p-charts' limits were determined according to an exact method based on the binomial distribution.¹⁵ Exact control and warning limits were set at 99.73% (3 SD from the mean) and 95.45% (2 SD from the mean), respectively (Appendix 1). The detection of a special cause variation was defined as a single point outside the control limits or two out of three successive points between a warning limit and a control limit on the same side of the central line.¹⁶

The agreement between crude and case-mix-adjusted control charts in detecting indicator variations was measured using the weighted Cohen κ statistic.¹⁷ The positions of the data points for both postoperative recurrent laryngeal nerve palsy and hypocalcaemia were compared in terms of five ordinal levels based on warning and control limits.

Statistical analyses were conducted using SPSS (release 12.0.0; SPSS, Chicago), and control charts were generated using Microsoft Office Excel 2007 (Microsoft Corporation, Redmond, Washington).

RESULTS

Study population

We evaluated the outcomes of 1405 thyroidectomies. Among these procedures, 1036 were total thyroidectomies, 195 were lobectomies, 97 were thyroidectomies with lymph node resection, and 77 were completion thyroidectomies. The makeup of the thyroid diseases that were operated on was 758 non-toxic multinodular goitres, 301 carcinomas, 176 non-toxic solitary nodules, 100 Graves disease and 70 other diagnoses. The median patient age was 53 years (range 9–93), and 78.7% of patients were women (1106/1405).

The overall proportions of immediate postoperative recurrent laryngeal nerve palsy and hypocalcaemia were 7.4% (103/1388) and 20.5% (243/1186), respectively. Among the 309 patients who had an immediate complication consecutive to bilateral thyroidectomy, 215 had hypocalcaemia, 67 had recurrent laryngeal nerve palsy, and 27 had both complications.

Crude versus case-mix-adjusted control charts

The type of thyroid surgery ($p = 0.001$) was associated with a risk of postoperative recurrent laryngeal nerve palsy, whereas patient gender ($p < 0.001$), age ($p < 0.001$) and the type of thyroid surgery ($p < 0.001$) were associated with a risk of postoperative hypocalcaemia (table 1).

There was one difference between the crude and case-mix-adjusted control charts related to postoperative recurrent laryngeal nerve palsy (figure 1): in May 2006, the indicator reached the adjusted warning limit but not the crude warning limit. There were two differences between the crude and case-mix-adjusted control charts related to the monitoring of postoperative hypocalcaemia (figure 2). In July 2007, the indicator just reached the control limit of the crude control chart, whereas it crossed over the control limit of the case-mix-adjusted control chart. In January 2006, the indicator reached the crude but not the adjusted warning limit. Investigation of the single special cause variation that was detected in July 2007 revealed two elements. First, operating-room renovations took place from July to September 2007, leading to a reduction in available operating time for the same number of patients (from 42 to 32 h per week). Second, one surgeon was away in July 2007, although the number of patients undergoing thyroid surgery remained constant, so in July almost all procedures were performed by the only other surgeon available.

Table 1 Factors associated with immediate postoperative complications

	Adjusted ORs † (95% CI)	
	Recurrent laryngeal nerve palsy (n = 1384)	Hypocalcaemia (n = 1182)
Patient gender		
Female	0.78 (0.48 to 1.25)	2.74 (1.74 to 4.34)
Male*	1.0	1.0
Patient age (years)		
Age ≤ 40	0.83 (0.47 to 1.46)	2.57 (1.70 to 3.88)
40 < age ≤ 50	0.72 (0.39 to 1.32)	2.03 (1.32 to 3.14)
50 < age ≤ 60	0.80 (0.46 to 1.39)	1.02 (0.65 to 1.61)
Age > 60*	1.0	1.0
Thyroid disease		
Malignant neoplasm	1.38 (0.47 to 4.08)	1.35 (0.55 to 3.30)
Non-toxic solitary nodule*	1.0	1.0
Non-toxic multinodular goitre	1.49 (0.54 to 4.12)	1.25 (0.53 to 2.95)
Graves disease	0.62 (0.13 to 2.83)	1.56 (0.60 to 4.08)
Others‡	0.86 (0.20 to 3.78)	1.05 (0.32 to 3.37)
Type of surgery		
Lobectomy	0.46 (0.15 to 1.36)	
Total thyroidectomy*	1.0	1.0
Completion thyroidectomy	1.50 (0.68 to 3.28)	1.20 (0.65 to 2.22)
Thyroidectomy with lymph-node resection	3.35 (1.67 to 6.75)	2.90 (1.69 to 4.97)

*Reference group.

† No interactions were observed in the models. The Hosmer–Lemeshow goodness-of-fit statistics for the recurrent laryngeal nerve palsy and hypocalcaemia models were 0.93 and 0.52, respectively. The c-statistics for the recurrent laryngeal nerve palsy and hypocalcaemia models were 0.65 (95% CI 0.60 to 0.71) and 0.68 (95% CI 0.64 to 0.72), respectively.

‡ Autoimmune thyroiditis and thyrotoxicosis with multinodular goitre or single thyroid nodule.

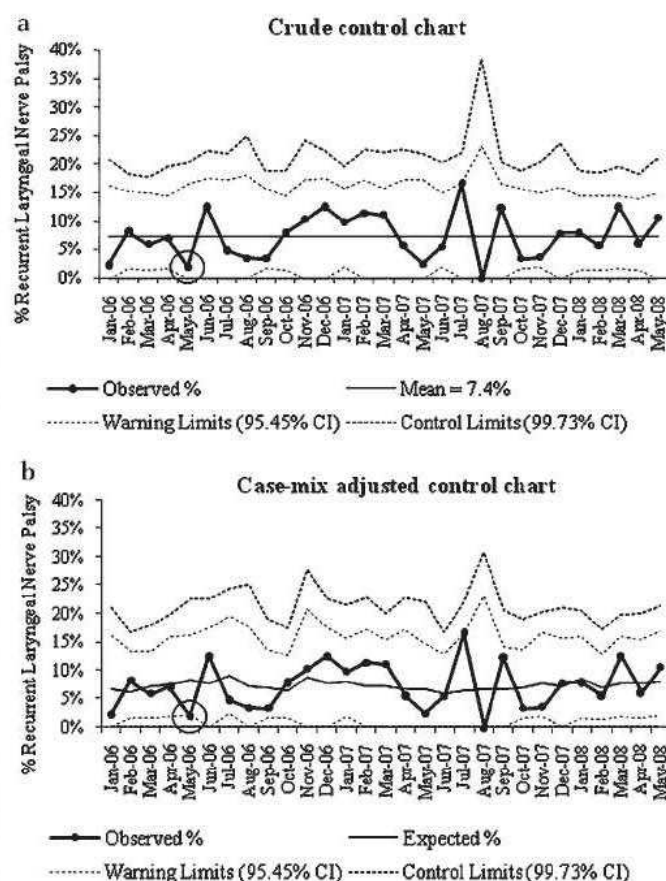


Figure 1 Crude versus case-mix adjusted p-control charts of immediate postoperative recurrent laryngeal nerve palsy.

The agreement between the crude and case-mix-adjusted control charts in detecting indicator variations (table 2) was 95% (95% CI 85% to 99%). The strength of agreement was $\kappa=0.76$ (95% CI 0.54 to 0.98).

DISCUSSION

There was good agreement between crude and case-mix-adjusted p-charts in detecting indicator variations for the monitoring of thyroid surgery safety. However, the single special cause of variation that occurred in July 2007 for hypocalcaemia monitoring was only detected by the case-mix-adjusted p-chart, suggesting that this method was slightly more sensitive than the crude control chart. Investigation of this cause revealed that the overactivity of a single surgeon was probably detrimental to the quality of thyroid surgery. Consequently, the primary concern should be to guarantee that surgeons can operate under satisfactory conditions and that they are not given excessive workloads.¹²

Performance monitoring using outcome indicators requires that the observed variations reflect true variations in patient safety. Variability in medical outcomes typically arises from a combination of four key elements: data quality, patient case-mix, quality of care and random effects.¹⁸ First, we assume that the observed variation in thyroid surgery outcomes is not influenced by the quality of data, since our monitoring system was implemented in a hospital where standardised modalities for detecting post-operative complications and for data collection were not changed during the study period. Second, provided that the adjustment was sufficient, and in view of the satisfactory agreement

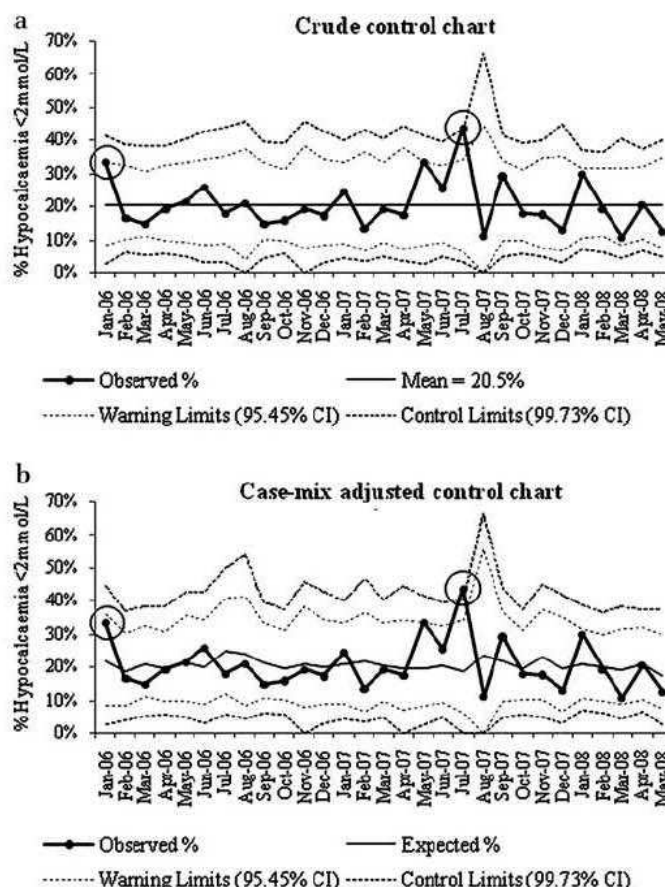


Figure 2 Crude versus case-mix adjusted p-control charts of immediate postoperative hypocalcaemia.

between crude and case-mix-adjusted p-charts, we also presume that the occurrence of complications was poorly influenced by the patient case-mix. This may be biased if adjustment is inadequate or if it does not include some unmeasured (eg, patient weight, large or thoracic goitre, and invasive cancer) or unknown case-mix factors, which might affect the outcome irrespective of quality of care.¹⁹ To assess whether the choice of variables for case-mix adjustment changed the conclusions of this study, we tested models based on different adjustment methods and observed similar results (data not shown). Third, the rationale for using case-mix-adjusted control charts is that the residual unexplained variation in outcomes and the detection of special causes are more likely to be attributable to the quality of care. The remaining variance is presumed to be due to the

Table 2 Agreement between crude and case-mix-adjusted p-charts in detecting indicator variations

Case-mix adjusted limits*		<WL	=WL	WL-CL	=CL	>CL	Total
Crude Limits*	<WL	53	1	0	0	0	54
	=WL	1	2	0	0	0	3
	WL-CL	0	0	0	0	0	0
	=CL	0	0	0	0	1	1
	>CL	0	0	0	0	0	0
	Total	54	3	0	0	1	58

*Each unit in the table represents the position of a data point on a recurrent laryngeal nerve palsy or hypocalcaemia control chart, according to five ordinal levels based on exact warning limits (WL, 95.45%) and control limits (CL, 99.73%), as follows: <WL (below the 95.45% limit), =WL (equal to the 95.45% limit), WL-CL (between the 95.45% and the 99.73% limits), =CL (equal to the 99.73% limit) and >CL (above the 99.73% limit).

characteristics of the individual surgeon (such as technical, physical and mental conditions) or to other organisational factors related to the surgical team (including equipment availability, team coordination or workload).²⁰ It would be useful to further adjust for variables regarding surgeons' compliance with evidence-based practices in order to restrict the sources of variability in medical outcomes to uncontrollable random effects. In thyroid surgery, adjustments could include the systematic visualisation¹¹ of recurrent laryngeal nerves in order to avoid injuries, as well as routine identification of at least two parathyroid glands in order to avoid permanent hypocalcaemia.²¹ Furthermore, to evaluate how one surgeon's performance influences the outcomes of the entire team, we could adjust for surgeon profile. Real-time monitoring of activity for each surgeon is also feasible using cumulative sum charts (CUSUM) to detect any defects in surgical processes as soon as possible.^{22 23} Focusing on individual performance allows each surgeon to check the quality of their daily practices and facilitates the identification of special cause variations. Nevertheless, compared with the simplicity of an adjusted p-chart, such an alternative tool can give less straightforward interpretations of variations in outcome indicators.

Though relatively simple and user-friendly, the value of the p-chart lies in its intuitive nature and the fact that the data are displayed graphically. Certainly, crude p-charts are fast and easy to produce, which provides quick performance feedback to the surgical staff in the context of real-time monitoring. One main limitation of these charts concerns the potential relationship between the occurrence of complications and case-mix characteristics. The clinical heterogeneity introduced by considering the outcomes of all procedures together in a broad grouping without case-mix adjustment is often argued to be clinically inappropriate. Due to the few differences observed between crude and case-mix-adjusted p-charts, our findings suggest that the expected gain from routine use of a case-mix-adjusted p-chart would be relatively minimal. Yet, as the great majority of data points were positioned between the lower and upper warning limits, statistical significance tests were likely to be underpowered and therefore a mere formality. Despite an acceptable fit to the data, the discriminatory power of the models was relatively poor. This might suggest that case-mix adjustment only controlled for a small part of the variability in thyroid surgery outcomes, which may simply demonstrate consistent care. This is perhaps not surprising, considering that thyroidectomy is a highly standardised procedure compared with other healthcare processes.²⁴ Additional confounding variables (such as other patient characteristics, surgeon profiles, organisational factors or adherence to evidence-based practices) could more precisely represent the variability in thyroid surgery outcomes and thus could be useful adjustments.

More research is needed to provide empirical evidence to corroborate the usefulness of the case-mix adjusted control chart in other settings.⁵ Although case-mix adjustment requires training in data analysis and is relatively time-consuming to integrate into the team's daily routine, it may be useful when indicator is positioned close to the limits of the crude control chart, which requires careful interpretation. The joint use of both crude and adjusted p-charts seems to be a reasonable approach to increase the accuracy of interpretation of outcome indicator variations in clinical practice.

Acknowledgements We are grateful to P Messy, for his helpful contribution to the study, and to A Favre, for the English revision of the manuscript.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Shewhart WA. *Economic control of quality of manufactured product*. New York: D Van Nostrand Company, 1931.
2. Mohammed MA, Cheng KK, Rouse A, et al. Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons. *Lancet* 2001;**357**:463–7.
3. Thor J, Lundberg J, Ask J, et al. Application of statistical process control in healthcare improvement: systematic review. *Qual Saf Health Care* 2007;**16**:387–99.
4. Iezzoni LI. Using risk-adjusted outcomes to assess clinical practice: an overview of issues pertaining to risk adjustment. *Ann Thorac Surg* 1994;**58**:1822–6.
5. Marshall T, Mohammed MA. Case-mix and the use of control charts in monitoring mortality rates after coronary artery bypass. *BMC Health Serv Res* 2007;**7**:63.
6. Lilford R, Mohammed MA, Spiegelhalter D, et al. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* 2004;**363**:1147–54.
7. Hart MK, Lee KY, Hart RF, et al. Application of attribute control charts to risk-adjusted data for monitoring and improving health care performance. *Qual Manag Health Care* 2003;**12**:5–19.
8. Alemi F, Oliver DW. Tutorial on risk-adjusted P-charts. *Qual Manag Health Care* 2001;**10**:1–9.
9. Thomusch O, Machens A, Sekulla C, et al. The impact of surgical technique on postoperative hypoparathyroidism in bilateral thyroid surgery: a multivariate analysis of 5846 consecutive patients. *Surgery* 2003;**133**:180–5.
10. Dralle H, Sekulla C, Haerting J, et al. Risk factors of paralysis and functional outcome after recurrent laryngeal nerve monitoring in thyroid surgery. *Surgery* 2004;**136**:1310–22.
11. Thomusch O, Machens A, Sekulla C, et al. Multivariate analysis of risk factors for postoperative complications in benign goiter surgery: prospective multicenter study in Germany. *World J Surg* 2000;**24**:1335–41.
12. Duclos A, Touzet S, Soardo P, et al. Patient safety monitoring in thyroid surgery using the Shewhart control chart. *Brit J Surg* 2009;**96**:171–4.
13. Hosmer DW, Lemeshow S. *Applied logistic regression*. New York: Wiley, 1989.
14. Aylin P, Bottle A, Majeed A. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. *BMJ* 2007;**334**:1044.
15. Mohammed MA, Worthington P, Woodall WH. Plotting basic control charts: tutorial notes for healthcare practitioners. *Qual Saf Health Care* 2008;**17**:137–45.
16. Benneyan JC, Lloyd RC, Plsek PE. Statistical process control as a tool for research and healthcare improvement. *Qual Saf Health Care* 2003;**12**:458–64.
17. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bulletin* 1968;**70**:213–20.
18. Mant J. Process versus outcome indicators in the assessment of quality of health care. *Int J Qual Health Care* 2001;**13**:475–80.
19. Pitches DW, Mohammed MA, Lilford RJ. What is the empirical evidence that hospitals with higher-risk adjusted mortality rates provide poorer quality care? A systematic review of the literature. *BMC Health Serv Res* 2007;**7**:91.
20. Vincent C, Moorthy K, Sarker SK, et al. Systems approaches to surgical quality and safety: from concept to measurement. *Ann Surg* 2004;**239**:475–82.
21. Pattou F, Combemale F, Fabre S, et al. Hypocalcemia following thyroid surgery: incidence and prediction of outcome. *World J Surg* 1998;**22**:718–24.
22. Woodall WH. The use of control charts in health-care and public-health surveillance. *J Qual Technol* 2006;**38**:89–104.
23. Rogers CA, Reeves BC, Caputo M, et al. Control chart for monitoring cardiac surgical performance and their interpretation. *J Thorac Cardiovasc Surg* 2004;**128**:811–19.
24. Hegner CF. A history of thyroid surgery. *Ann Surg* 1932;**95**:481–92.

APPENDIX 1: METHODOLOGY FOR THE CONSTRUCTION OF BINOMIAL EXACT LIMITS

According to the normal distribution, after rounding values, 95.45% of all plotted data are expected to fall within 2 SD, with the remaining 4.55% falling more than 2 SD away from the mean. Similarly, 99.73% of all plotted data are expected to fall within 3 SD of the mean if the process is stable (common cause variations), with the remaining 0.27% falling more than 3 SD away from the mean (special cause variations).

We used the cumulative binomial distribution to find the lower and upper control and warning limits according to the number of patients each month and the overall proportion of events for the crude p-chart, or the expected proportion of events each month for the case-mix adjusted p-chart.

For the crude chart, we used:

$$F_{\text{crude}}(x_t, \bar{p}, n_t) = P_{\text{crude}}(X_t \leq x_t) = \sum_{i=0}^{x_t} \binom{n_t}{i} \bar{p}^i (1 - \bar{p})^{n_t - i},$$

where \bar{p} is the overall observed proportion of complications, and n_t is the number of patients each month.

SD	Percentage
1 SD	68.2689492%
1.645 SD	90%
1.960 SD	95%
2 SD	95.4499736%
2.576 SD	99%
3 SD	99.7300204%
3.2906 SD	99.9%
4 SD	99.993666%
5 SD	99.99994267%
6 SD	99.999998027%
7 SD	99.999999997440%

For the adjusted chart, we used:

$$F_{\text{adjusted}}(x_t, p_t, n_t) = P_{\text{adjusted}}(X_t \leq x_t) = \sum_{i=0}^{x_t} \binom{n_t}{i} p_t^i (1 - p_t)^{n_t-i},$$

where p_t is the expected proportion of complications each month, and n_t is the number of patients each month.

The lower warning limit for each month is the value of x_t that satisfies $P(X_t \leq x_t) = \frac{1-0.9545}{2} = 0.02275$, and the upper warning limit is the value of x_t that satisfies $P(X_t \leq x_t) = \frac{1+0.9545}{2} = 0.97725$ for 2 DS (95.45%).

The lower control limit for each month is the value of x_t that satisfies $P(X_t \leq x_t) = \frac{1-0.9973}{2} = 0.00135$, and the upper control limit is the value of x_t that satisfies $P(X_t \leq x_t) = \frac{1+0.9973}{2} = 0.99865$ for 3 DS (99.73%).

In practice, we used the CRITBINOM function of MS Excel.

5.3. Étude de l'effet du neuromonitoring sur la performance du chirurgien et courbe d'apprentissage liée à son utilisation par carte CUSUM (*article 4*)

Le neuromonitoring représente un apport intéressant en chirurgie thyroïdienne pour faciliter le repérage des nerfs récurrents et ainsi éviter leur éventuelle lésion au moment de la résection unilatérale ou bilatérale de la glande. Néanmoins, son intérêt réel à prévenir la survenue d'une paralysie récurrentielle est controversé et largement débattu aujourd'hui entre les chirurgiens. Peu d'études scientifiques rigoureuses ont évalué son efficacité et leurs résultats s'avèrent contradictoires, peut-être en raison de l'absence de prise en compte de la courbe d'apprentissage liée à la maîtrise du neuromonitoring par le chirurgien. De plus, les modalités d'utilisation de cet outil demeurent encore peu standardisées.

Cette troisième partie du projet visait à investiguer l'effet du neuromonitoring sur la performance individuelle du chirurgien en tentant d'élucider deux questions complémentaires : une utilisation routinière du neuromonitoring influence-t-elle la technique du chirurgien pour identifier et disséquer les nerfs récurrents ? Combien de thyroïdectomies un chirurgien doit-il réaliser avant de maîtriser cet outil ?

Pour chacun des 3 chirurgiens du service, nous avons tracé la courbe d'apprentissage relative aux 475 chirurgies thyroïdiennes réalisées de mai 2008 à avril 2009 en utilisant le neuromonitoring. La carte O-E et le test LC-CUSUM ont montré que la durée d'apprentissage variait fortement d'un chirurgien à l'autre, allant de 35 à 304 procédures pour une maîtrise complète de cet outil. De telles différences pouvaient refléter un degré de remodelage variable de la technique chirurgicale propre à chacun d'entre eux. Le chirurgien le plus expérimenté présentait la durée d'apprentissage la plus longue, en raison d'une modification importante de son geste opératoire. Le neuromonitoring pouvait être utilisé pour réaliser une chirurgie moins invasive, sans dissection complète des nerfs récurrents. Cette technique nécessitait de l'entraînement et du temps pour être maîtrisée.

Antoine Duclos, Jean-Christophe Lifante, Simon Ducarroz, Pietro Soardo, Sandrine Touzet, Cyrille Colin, Jean-Louis Peix.

Influence of intraoperative neuromonitoring on surgeon's technique during thyroidectomy
(article soumis).

Le quatrième article soumis et reproduit ci-après présente une évaluation de l'effet du neuromonitoring en pratique courante sur la performance du chirurgien, incluant l'analyse de la courbe d'apprentissage liée à son utilisation. Cet exemple, appliqué à la chirurgie thyroïdienne, illustre l'influence que peut avoir une technologie sur la technique opératoire pour parvenir à une chirurgie plus sûre. Il montre la nécessité de tenir compte des effets d'apprentissage individuels lorsque l'on souhaite mettre en évidence l'efficacité d'une nouvelle technique chirurgicale.

Influence of intraoperative neuromonitoring on surgeon's technique during thyroidectomy

Antoine Duclos, M.D.,¹ Jean-Christophe Lifante, M.D., Ph.D.,² Simon Ducarroz, M.Sc.,¹ Pietro Soardo, M.D.,² Sandrine Touzet, M.D.,¹ Cyrille Colin, M.D., Ph.D.,¹ Jean-Louis Peix, M.D., Ph.D.²

¹ Hospices Civils de Lyon, Pôle Information Médicale Evaluation Recherche, Lyon, F-69003, France; Université de Lyon, EA Santé-Individu-Société 4129, Lyon, F-69002 France.

² Hospices Civils de Lyon, Centre Hospitalier Lyon Sud, Service de Chirurgie Générale et Endocrinienne, Pierre Bénite, F-69495, France.

Abstract

Background: When assessing the value of intraoperative nerve monitoring (IONM) during routine thyroidectomy, it is necessary to consider its influence on the surgeon's dissection technique. We investigated the effect of IONM on the individual performance of surgeon by determining the learning curve associated with this tool.

Methods: A one year prospective study was performed between May 2008 and April 2009 within a team of three experienced endocrine surgeons. The measurement of surgical performance was based on the detection of immediate postoperative recurrent laryngeal nerve palsy using laryngoscopy. Individual learning curves associated with IONM acquisition were drawn using the cumulative sum charts (CUSUM). Each surgeon was questioned about the possible changes he had experienced in his own surgical technique following the introduction of IONM.

Results: A total of 475 consecutive patients who underwent thyroid surgery using IONM were included. The pattern of learning curves varied among surgeons and ranged from 35 to 304 procedures required for complete IONM acquisition. The surgeon with the longest learning curve also described a deep modification in his technique related to nerve dissection.

Conclusions: IONM can induce changes in surgical practice to improve nerve preservation. The different patterns of learning curve among surgeons may reflect the variable degree of remodeling in their own dissection technique. Such learning effect must be considered when assessing the impact of using IONM on patient safety.

Introduction

Injury of the recurrent laryngeal nerve (RLN) constitutes one of the most serious complications that can occur during thyroidectomy, and is the leading cause of medicolegal litigation after endocrine surgery.^{1,2} This injury is responsible for vocal fold paralysis, which is associated with dysphonia, difficulty swallowing, and respiratory problems.³

The routine identification of nerves during thyroid surgery has been advocated since the 1930s⁴ and has significantly reduced the RLN palsy rate.^{5,6} Nevertheless, the rates of temporary and permanent RLN palsy after thyroidectomy remain high at 9.8% and 2.3%, respectively, when a systematic rigorous postoperative laryngoscopy is performed.⁷ The intraoperative nerve monitoring (IONM) has been proposed to help trace the anatomical

pathways of nerves and prevent injuries,⁸ but it remains controversial among surgeons.⁹ Considering the divergent results obtained from cases series and nonrandomized trials, the addition of IONM does not appear to further reduce the incidence of postoperative nerve palsy and its current use remains poorly standardized.¹⁰

Before promoting the widespread implementation of IONM into routine practice, it is necessary to better understand the influence of this tool on surgeon's performance.¹¹ Does the surgical technique for identifying nerves change following the introduction of IONM? And how many thyroidectomies must a surgeon perform before mastering this tool? The value of IONM may depend on the surgeon's previous background and technical skill regarding the visualization and dissection of nerves, but also on the technical adaptation induced by this

tool. We aimed to investigate the effect of IONM on the surgeon's performance by determining the individual learning curve associated with this tool.

Methods

Study population and data collection

A prospective study was conducted between May 2008 and April 2009 in the department of endocrine surgery at a university hospital referral centre. The surgical team under observation consisted of three experienced surgeons, including one very high volume surgeon (surgeon A had performed more than 300 thyroidectomies a year for 25 years) and two high volume surgeons (surgeons B and C had each performed more than 150 thyroidectomies a year for 5 years), none of whom had ever used IONM before the beginning of the study.¹²

All patients who underwent a thyroid surgery using neuromonitoring during the 12-month period were included. They had IONM performed as part of their routine surgical course and the study was undertaken with the understanding and informed consent of each. Postoperative vocal fold mobility was systematically assessed by anesthetists using fiberoptic transnasal laryngoscopy in recovery room after a complete waking up of the patient. When this was inconclusive, the patient was not fully cooperative or the voice was abnormal, laryngoscopy was repeated within 48 hours by an ENT specialist. In case of patient with abnormal voice or having had a previous thyroidectomy, a laryngoscopy was performed preoperatively. Accordingly, RLN palsy was accounted only for those patients with new postoperative palsy.

Specific information about each thyroid surgery (detailed procedure, utilization of IONM and surgeon identity) and its related complications (in-hospital transient RLN palsy) was collected prospectively using a standardized form. Other information was obtained from the hospital information system, which contains standard discharge abstracts including compulsory information on the patients' demographics, the diagnoses of operated thyroid diseases and the length of stay. After the study period, each surgeon was separately questioned about the potential changes he had experienced in his own

surgical technique due to the introduction of IONM.

Surgical technique and IONM utilization

Before the introduction of neuromonitoring in May 2008, all surgeons performed the thyroidectomies with the same operative plan. The recurrent nerve was rigorously visualized and completely dissected, using the inferior thyroid artery as a landmark after the ligation of the upper thyroid vessels and before the section of the inferior thyroid vessels.

Due to the acquisition of a single neurominotoring device for the entire surgical department, we could not use this technique for all patients who underwent a thyroidectomy during the study period. The surgeons did not select patients and always used IONM when available. During the induction of anesthesia, if the surgeon had elected to use IONM, the anesthesiologist set up endotracheal tube electrodes prior to intubation. The vocal cords were systematically observed to be in contact with the left and right endotracheal tube electrodes. The NIM-Response Nerve Integrity Monitoring System (Xomed Medtronic, Jacksonville, FL, USA) was used in all procedures with a supramaximum stimulation set at 2 mA. The IONM was used either to confirm the path of the nerve after its visualization or to help locate it in cases where identification was difficult. A positive signal was determined by observing the electromyogram. The vagus and recurrent laryngeal nerves were stimulated on both sides before and after resection, and in case of loss of signal at the time of the lesion occurred in order to differentiate between artifacts and functional nerve dysfunctions.

Cumulative sum charts characteristics

The individual surgeons' progress in mastering the IONM procedure was monitored based on the cumulative occurrence of success (no complication) or failure (immediate postoperative RLN palsy) for each surgeon. Related learning curves were drawn using two complementary graphs as follows: the cumulative observed minus expected failure chart (O-E),¹³ which gives an intuitive visual interpretation of the learning curve, and the cumulative summation test for learning curves (LC-CUSUM),¹⁴ which provides a formal

statistical test to detect when the new skill is acquired by the surgeon.

The O-E chart showed whether a surgeon's performance was above or below that which could be expected by plotting the difference between the actual and expected number of complications against the operation number in chronological sequence. For each procedure, there was an expected risk of complication that determined the magnitude by which the graph ascended or descended. For a success, the plot decreased by an amount equal to the estimated probability of complication. For a failure, the plot increased by an amount equal to the estimated probability of a procedure without complication.¹³

The LC-CUSUM was specifically developed to determine when the learning curve of a surgeon is complete by reaching a predefined level of performance. A lower decision limit was drawn horizontally to determine at what time the surgeon has the process under control. In this study the acceptable failure rate was set at 7.5% (ie, the baseline prevalence of RLN palsy for the surgical team)¹⁵ and the unacceptable failure rate was set at 14%. The lower decision limit was positioned at -1.35 to yield average run length (ARL) values of 119 and 34 for ARL₀ and ARL₁, respectively, signifying that a false alarm would occur every 119 procedures for a process that was out of control and that 34 procedures would be necessary to trigger an alarm if the process was in control.¹⁴

Case-mix adjustment

Case-mix adjustment was based on a set of relevant variables reported to be associated with complications in the literature, including the patients' gender, age, thyroid disease diagnosis, and type of thyroidectomy.^{16,17} Odds-ratios were estimated from data obtained before IONM introduction (from January 2006 to April 2008) using a multivariable logistic regression.¹⁵ For each thyroid procedure, the probability of RLN palsy was computed from the model estimates, according to the patient's preoperative individual risk.

Case-mix adjustment was performed using SPSS® (SPSS Inc., release 12.0.0, Chicago, IL, USA) and simulations to determine the ARLs were computed using R® software (R

Foundation for Statistical Computing, Vienna, Austria).

Results

Between May 2008 and April 2009, 475 thyroid surgeries were performed using IONM (Table 1). There were 36 cases of immediate postoperative RLN palsy, accounting for 7.6% of interventions (95% Confidence Interval = 5.3%-10.1%). Most procedures were performed by surgeon A (72.4%), followed by surgeons B (19.8%) and C (7.8%).

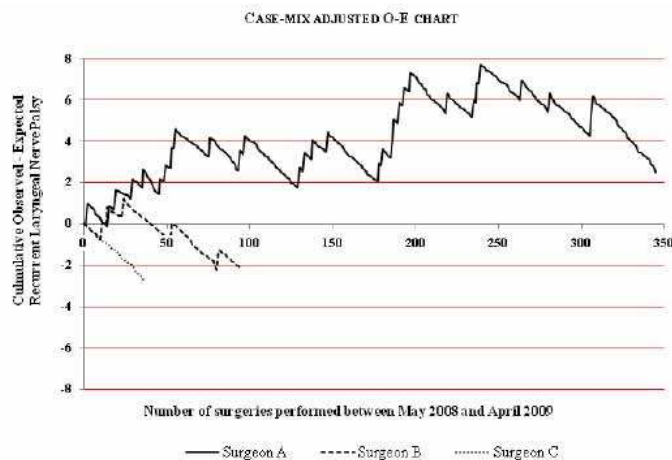
Table 1. Thyroid procedures characteristics

	Thyroid procedures (n = 475)
Patient demographics	
Women, No. (%)	361 (77.0)
Age, median (range), y	53.3 (11.7-83.2)
Thyroid disease, No. (%)	
Malignant neoplasm	108 (22.7)
Non-toxic solitary nodule	28 (5.9)
Non-toxic multinodular goiter	245 (51.6)
Grave's disease	40 (8.4)
Other*	54 (11.4)
Type of surgery, No. (%)	
Lobectomy	32 (6.7)
Total thyroidectomy	360 (75.8)
Completion thyroidectomy	40 (8.4)
Thyroidectomy with lymph node dissection	43 (9.1)
Postoperative outcome	
Recurrent laryngeal nerve palsy, No. (%)	36 (7.6)
Length of stay, mean (SD), d	2.7 (1.0)
First operator, No. (%)	
Surgeon A	344 (72.4)
Surgeon B	94 (19.8)
Surgeon C	37 (7.8)

* Autoimmune thyroiditis, thyrotoxicosis with a multinodular goiter or a single thyroid nodule.

Analysis of the O-E chart (Figure 1) for the whole series of surgeries revealed that the surgeon A had 2.5 more complications than expected. Two short series from the 15th to 55th and from the 177th to 197th procedures were initially distinguished as having a poor performance before a high performance series occurred from the 239th to the 344th procedure. Surgeon B had 2.1 fewer complications than expected, and a high performance series was observed from the 25th to the 94th procedure. Surgeon C had 2.8 fewer complications than expected, and had a high performance series that began at the start of monitoring.

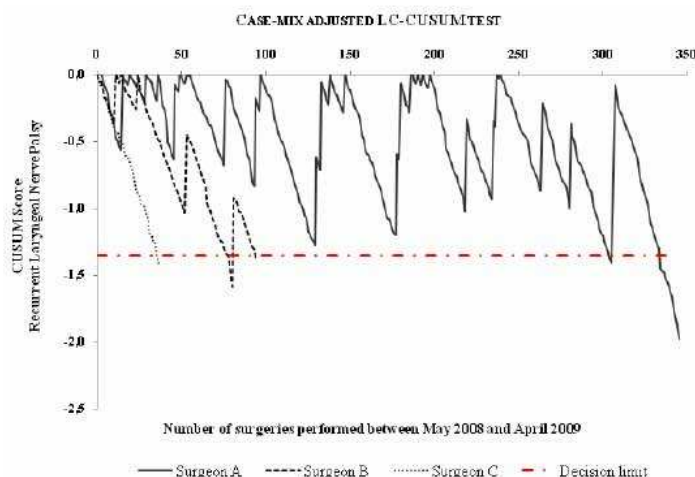
Figure 1. Case-mix adjusted cumulative observed minus expected graph of postoperative recurrent laryngeal nerve palsy, stratified by surgeon.



The graph moved upward if the failure rate increased above that predicted by the risk model, it moved downward if the rate decreased, and it oscillated around the horizontal axis if the performance was in line with the predicted risks.

The LC-CUSUM (Figure 2) for surgeons A, B and C first crossed the decision limit by the 303rd, the 78th and the 35th procedures, respectively.

Figure 2. Case-mix adjusted cumulative summation test for the learning curve of postoperative recurrent laryngeal nerve palsy, stratified by surgeon.



When the graph crossed the decision limit, this signaled that sufficient evidence had accumulated to indicate that the learning curve of a surgeon was complete.

All surgeons claimed that the neuromonitoring improved the feeling of safety during thyroid surgery but have described markedly different approaches for the nerve dissection. Surgeons B and C did not changed their surgical

technique and they kept visualizing and dissecting the RLN the same manner as before the implementation of IONM. Conversely, the most experienced surgeon reported substantial changes regarding the nerve identification. Since he started using IONM, the surgeon A did not perform systematically a complete dissection of the nerve. He gradually performed a less invasive dissection by checking the RLN location with the IONM without systematic visualization and by performing the resection of the thyroid gland very close to the capsule.

Discussion

We prospectively followed 475 patients who underwent thyroid surgery using the neuromonitoring. The changes in the outcomes of a team of high volume surgeons were individually monitored over time. The pattern of the learning curve associated with IONM acquisition varied by surgeons and may reflect changes in their own dissection technique. On one hand, the two surgeons with the lowest experience required fewer than 100 procedures before mastering the tool. They kept on performing the same rigorous dissection, using the IONM for not confusing the nerve with blood vessels. On the other hand, following an initial worsening, the most experienced surgeon finally achieved an acceptable level of performance after 300 thyroid procedures, once having deeply questioned his practice. Assuming that a too invasive dissection could be responsible for RLN palsies, he strived to change his surgical technique for better outcomes. He used the IONM to perform thyroidectomy without completely dissecting the nerve. This innovative technique needed a longer duration of IONM's learning than the conventional technique.

Previous studies have also reported a variable duration of the learning curve associated with the routine introduction of a safe IONM technique.^{18,19} This may be related to the differences between surgeons regarding their usual operative techniques (such as routine nerves visualization and careful dissection) and their ability to incorporate a new skill into practice. Although the success of a thyroidectomy is partially determined by the surgeon's experience¹⁷ and his stage of training,²⁰ we speculate that surgeons would have less difficulty learning new techniques if

they changed their surgical habits at the beginning of their career.²¹

One danger of this type of uncontrolled study design is confounding by secular trends. Standardized modalities for detecting RLN palsy and data collection did not change over time, and the occurrence of complications was adjusted for patient case-mix. Accordingly, we presume that the observed variation in thyroid surgery outcomes was mostly attributable to the implementation of IONM and progressive acquisition of the skills necessary to adequately perform it. Nevertheless, we cannot formally exclude the possibility that unknown changes within the surgical team may have occurred and influenced surgical outcomes.²² Furthermore, the rate of RLN palsy may seem high because we did not excluded patients requiring nerve resection in the case of infiltrated carcinoma and we have objectively screened the immediate postoperative complications using systematic direct laryngoscopy. Considering the frequent resolution of these transient complications with time¹⁷ and the restricted number of surgeons studied, our findings should be confirmed in other care facilities by monitoring the rate of permanent palsy in terms of nerves at risk.

Most of the controlled studies on this topic have failed to provide evidence that IONM is effective at preventing RLN palsy. Two explanations can typically be argued. The first one is methodological and refers to inadequate statistical power¹⁰ or to the incomplete learning of IONM by surgeons due to an insufficient number of interventions. Indeed, learning effects may bias results of clinical trials because the performance of surgeons is expected to improve over time within the IONM arm, while remaining stable for controls.²³ The second explanation just advocates that IONM does not replace the gold standard of rigorous nerves identification. As a routine adjunct, IONM would not further improve the surgical safety, when surgeons got used to perform a systematic nerves visualization and dissection. Only one recent randomized trial, which was conducted in a single centre, found that nerve monitoring provided an advantage in avoiding transient RLN palsy compared with visualization alone.²⁴ The reproducibility of these findings should now be tested within a representative

sample of surgeons by conducting a large cluster-randomized trial. Given our findings, participating surgeons should previously be familiar with neuromonitoring (ie, having completed their learning curve),²⁵ or statistical analysis should consider IONM skill acquisition during the course of the study.²⁶

From the surgeon's perspective, neuromonitoring provides comfort during thyroidectomy for identification, dissection and control of recurrent nerves. Even if it can lead to a false sense of security, IONM can reduce the stress associated with nerve dissection when a surgeon operates on a patient with a challenging anatomy or in cases of extensive surgery.⁹ IONM induces technical changes in surgical practice for preserving the recurrent nerves, consciously or not. It can be used to perform a less invasive thyroid resection without complete nerves dissection. Although this technique requires training and time to be safe, IONM helps the surgeon in mastering it. Accordingly, the different patterns of learning curve among surgeons may reflect the variable degree of remodeling in their own dissection technique. Such learning effect must be considered by researchers when conducting comparative effectiveness on a new surgical technique to improve patient safety.

Acknowledgements

The authors are grateful to David J Biau (Service de Chirurgie Orthopédique et Traumatologique, Hôpital Cochin, Paris, France) for his helpful contributions in designing the LC-CUSUM and to Philippe Messy for extracting the study data from the hospital information system.

Conflict of interest & funding support

The authors declare no conflict of interest or funding support.

References

1. Kern KA. Medicolegal analysis of errors in diagnosis and treatment of surgical endocrine disease. *Surgery* 1993;114(6):1167-73.
2. Shaw GY, Pierce E. Malpractice litigation involving iatrogenic surgical vocal fold paralysis: a closed-claims review with recommendations for prevention and management. *Ann Otol Rhinol Laryngol* 2009;118(1):6-12.
3. Munch S, de Kryger L. A piece of my mind. Moral wounds: complicated complications. *JAMA* 2001;285(9):1131-2.
4. Lahey FH, Hoover WB. Injuries to the recurrent laryngeal nerve in thyroid operations: their management and avoidance. *Ann Surg* 1938;108(4):545-62.
5. Jatzko GR, Lisborg PH, Müller MG, Wette VM. Recurrent nerve palsy after thyroid operations--principal nerve identification and a literature review. *Surgery* 1994;115(2):139-44.
6. Hermann M, Alk G, Roka R, Glaser K, Freissmuth M. Laryngeal recurrent nerve injury in surgery for benign thyroid diseases: effect of nerve dissection and impact of individual surgeon in more than 27,000 nerves at risk. *Ann Surg* 2002;235(2):261-8.
7. Jeannon JP, Orabi AA, Bruch GA, Abdalsalam HA, Simo R. Diagnosis of recurrent laryngeal nerve palsy after thyroidectomy: a systematic review. *Int J Clin Pract* 2009;63(4):624-9.
8. Shedd DP, Burget GC. Identification of the recurrent laryngeal nerve. *Arch Surg* 1966;92(6):861-4.
9. Sturgeon C, Sturgeon T, Angelos P. Neuromonitoring in thyroid surgery: attitudes, usage patterns, and predictors of use among endocrine surgeons. *World J Surg* 2009;33(3):417-25.
10. Dralle H, Sekulla C, Lorenz K, Brauckhoff M, Machens A; German IONM Study Group. Intraoperative monitoring of the recurrent laryngeal nerve in thyroid surgery. *World J Surg* 2008;32(7):1358-66.
11. Haynes B. Can it work? Does it work? Is it worth it? The testing of healthcare interventions is evolving. *BMJ* 1999;319(7211):652-3.
12. Duclos A, Touzet S, Soardo P, Colin C, Peix JL, Lifante JC. Quality monitoring in thyroid surgery using the Shewhart control chart. *Brit J Surg* 2009;96(2):171-4.
13. Rogers CA, Reeves BC, Caputo M, et al. Control chart for monitoring cardiac surgical performance and their interpretation. *J Thorac Cardiovasc Surg* 2004;128:811-9.
14. Biau DJ, Williams SM, Schlup MM, Nizard RS, Porcher R. Quantitative and individualized assessment of the learning curve using LC-CUSUM. *Br J Surg* 2008;95(7):925-9.
15. Duclos A, Voirin N, Touzet S, Soardo P, Schott AM, Colin C, et al. Crude versus case-mix-adjusted control charts for safety monitoring in thyroid surgery. *Qual Saf Health Care* 2009; [in press].
16. Dralle H, Sekulla C, Haerting J, Ganesh JS, Bonser RS, Angelini GD. Risk factors of paralysis and functional outcome after recurrent laryngeal nerve monitoring in thyroid surgery. *Surgery* 2004;136(6):1310-22.
17. Thomusch O, Machens A, Sekulla C, et al. Multivariate analysis of risk factors for postoperative complications in benign goiter surgery: prospective multicenter study in Germany. *World J Surg* 2000;24(11):1335-41.
18. Dionigi G, Bacuzzi A, Boni L, Rovera F, Dionigi R. What is the learning curve for intraoperative neuromonitoring in thyroid surgery? *Int J Surg* 2008;6 Suppl 1:S7-12.
19. Jonas J, Bähr R. Intraoperative neuromonitoring of the recurrent laryngeal nerve - results and learning curve. *Zentralbl Chir* 2006;131(6):443-8.
20. Lamadé W, Renz K, Willeke F, Klar E, Herfarth C. Effect of training on the incidence of nerve damage in thyroid surgery. *Br J Surg* 1999;86(3):388-91.
21. Reznick RK, MacRae H. Teaching surgical skills - changes in the wind. *N Engl J Med* 2006;355(25):2664-9.
22. Vincent C, Moorthy K, Sarker SK, Chang A, Darzi AW. Systems approaches to surgical quality and safety: from concept to measurement. *Ann Surg* 2004;239(4):475-82.
23. Lilford R, Braunholtz D, Harris J, Gill T. Trials in surgery. *Br J Surg* 2004;91(1):6-16.
24. Barczyński M, Konturek A, Cichon S. Randomized clinical trial of visualization versus neuromonitoring of recurrent laryngeal nerves during thyroidectomy. *Br J Surg* 2009;96(3):240-6.
25. Paradis C. Bias in surgical research. *Ann Surg* 2008;248(2):180-8.
26. Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT. Statistical assessment of the learning curves of health technologies. *Health Technol Assess* 2001;5(12):1-79.

6. Synthèse

6.1. Discussion des travaux réalisés

6.1.1. Intérêt de la carte de contrôle

Principaux enseignements appris de l'application des cartes à la chirurgie thyroïdienne

Ce travail restitue les principaux résultats de la mise en œuvre d'un suivi d'indicateurs dans un service présentant une activité importante de chirurgie thyroïdienne. Pendant près de quatre ans, nous avons analysé la performance collective et individuelle d'une équipe de chirurgiens pour en comprendre les variations au cours du temps. Cette expérience a montré l'applicabilité des cartes de contrôle à la chirurgie de la thyroïde ainsi que leur intérêt pour aider les chirurgiens à mieux interpréter leurs résultats et à mieux maîtriser les facteurs y concourant.

Le premier enseignement concerne la faisabilité de notre démarche qui s'est avérée relativement simple à mettre en œuvre malgré l'absence de budget de recherche spécifique. Le soutien du chef de service de même que l'instauration d'une atmosphère d'évaluation constructive et la volonté des chirurgiens d'évaluer leurs résultats en ont facilité l'acceptabilité, se traduisant par leur participation active au protocole d'évaluation.¹²² La logistique nécessaire à son implémentation sur le terrain était particulièrement adaptée aux habitudes de l'équipe, puisque la construction des indicateurs reposait sur les complications déjà dépistées en routine et sur la collecte usuelle des données dans le PMSI. Le surcoût de l'évaluation était donc modeste, imputable principalement au codage supplémentaire des complications de la chirurgie thyroïdienne, au développement initial des cartes pour automatiser leur production et au temps humain consacré aux réunions de restitution trimestrielles.

Le second enseignement porte sur l'utilité des cartes de contrôle et les nombreux bénéfices qui ont été observés suite à leur introduction. En facilitant l'interprétation des variations observées sur les indicateurs, elles ont permis de générer de la connaissance sur les facteurs à l'origine de ces variations et ont clairement favorisé l'émergence d'hypothèses de recherche.¹⁰⁶ Elles ont également incité les chirurgiens à garder continuellement un regard critique sur leurs pratiques et à mettre à l'épreuve leurs idées au service d'une amélioration de la sécurité du patient.

L'essentiel des efforts à fournir pour améliorer la sécurité du patient opéré se situait en amont et en aval de la mesure. Le succès de l'introduction d'une carte en routine reposait principalement sur une bonne connaissance de la prise en charge évaluée, sur l'investigation rigoureuse des causes spéciales de variations et sur l'expérimentation d'actions concrètes destinées à améliorer la qualité de la chirurgie (table XV). La carte de contrôle représentait un prétexte pour s'y intéresser en offrant aux chirurgiens la possibilité d'évaluer leur pratique et de s'interroger sur comment faire mieux. L'expérience accumulée, après deux années de suivi régulier des indicateurs, nous a ainsi encouragés à recentrer le contenu des réunions de restitution trimestrielles sur la compréhension des variations d'indicateurs observées et sur la préparation d'actions d'améliorations spécifiques. Avant d'envisager une réduction des complications, il était nécessaire de fiabiliser le processus chirurgical, c'est-à-dire d'en réduire la variabilité par l'élimination de toutes les cause spéciales détectées. Une fois le processus maîtrisé, aucune amélioration ne pouvait survenir spontanément sans la conception et l'introduction d'actions ciblées sur la prise en charge. Notre approche s'est naturellement orientée vers le modèle de production de connaissance développé par Shewhart, à savoir : concevoir une action d'amélioration, la mettre en œuvre puis en évaluer l'effet.⁹⁵ Cela impliquait que chaque action soit testée de manière isolée afin d'en évaluer l'impact indépendamment d'autres changements apportés à la prise en charge.

Table XV. Recommandations pour l'implémentation d'une carte de contrôle en routine

1.	Désignation d'un chef de projet motivé au sein de l'équipe placée sous surveillance
2.	Choix du processus de soins à mettre sous contrôle
3.	Fixation d'objectifs d'amélioration à atteindre
4.	Sélection pragmatique des indicateurs à suivre et modalités de construction
5.	Choix du mode de recueil et d'extraction informatisé des données
6.	Choix du type de carte de contrôle utilisé, du paramétrage et des règles d'interprétation
7.	Choix des méthodes d'identification des causes spéciales
8.	Proposition d'actions d'amélioration des soins concrètes à tester
9.	Planification de la périodicité de restitution des cartes dans l'organisation
10.	Établissement du calendrier et répartition des tâches entre les personnes impliquées

Des constatations analogues aux nôtres ont été rapportées dans le cadre d'expériences proches, reposant sur la mise en œuvre d'un contrôle qualité prospectif au bloc opératoire. Menées à l'échelle d'un service ou d'un hôpital et s'intéressant à la performance individuelle de l'opérateur ou de l'ensemble de l'équipe chirurgicale, elles ont montré l'utilité des cartes

dans diverses spécialités chirurgicales.^{96,117} Des cartes de contrôle ont ainsi été testées avec succès en anesthésie,^{216,217,218} en chirurgie cardio-thoracique,^{219,220} en orthopédie,²²¹ en urologie²²² ou en chirurgie pédiatrique.²²³

Bien qu'il puisse exister des barrières à son application, la carte de contrôle représente un outil puissant, dont l'emploi peut aider les équipes à apprendre de leurs prises en charge. Parmi ses principales limites d'utilisation, on retiendra que l'identification des causes spéciales de variations n'est pas toujours évidente, que le profil des patients opérés doit être considéré attentivement, de même que la qualité des données utilisées ou les moyens nécessaires à leur collecte et à leur analyse. Trouver le meilleur équilibre possible entre le niveau d'agrégation des données et la périodicité de restitution des cartes peut également s'avérer délicat. Afin d'autoriser un suivi en temps réel des indicateurs, la mise à disposition d'un système d'information alimenté en routine et informatisé est essentielle. Enfin, l'actualisation et la restitution régulières des cartes aux équipes, ainsi que l'accompagnement de leur interprétation par des experts en contrôle qualité, constituent des facteurs clés de réussite de la démarche.^{96,105}

Une meilleure maîtrise des résultats de la chirurgie thyroïdienne

L'application des cartes s'est immédiatement accompagnée d'une réduction importante et pérenne des hypocalcémies postopératoires (passant de 32% à 21%). Nous avons attribué cette amélioration à la plus grande vigilance des chirurgiens du fait de leur mise sous observation (effet Hawthorne)⁹⁴ et des fréquents retours d'informations sur leur performance (*audit-feedback*).²²⁴

Lors du suivi prospectif, l'utilisation des cartes de contrôle a conduit à une meilleure maîtrise des résultats de la chirurgie thyroïdienne mais n'a pas permis d'améliorer significativement le niveau de performance de l'équipe chirurgicale. Réduire progressivement la survenue des complications opératoires s'est révélé être un challenge particulièrement difficile et la solution pour y parvenir ne pouvait être espérée du seul suivi des indicateurs.²²⁵

L'application formelle des cartes au contrôle qualité pour guider l'interprétation des indicateurs a révélé une détérioration de la performance collective des chirurgiens au cours de l'été 2007 (Figure 14). Un changement temporaire de bloc opératoire et une répartition inadéquate de leurs congés annuels ont été identifiés comme causes spéciales de cette variation. Malgré une diminution des plages opératoires mises à disposition sur la période, le nombre de thyroïdectomies réalisées est resté stable, entraînant une suractivité chirurgicale

temporaire. Le maintien d'un planning opératoire adapté au volume d'activité des chirurgiens et un meilleur lissage de leur charge de travail ont donc été préconisés pour prévenir la récurrence d'une telle situation.

Depuis, aucun dépassement des limites de contrôle n'a été détecté et nous avons observé une stabilisation des résultats de la chirurgie avec une tendance à la diminution progressive des paralysies récurrentielles suite à l'introduction du neuromonitoring. Le tracé des courbes d'apprentissage individuelles a révélé un délai d'acquisition variable de cette technologie, suggérant que l'expérience et la technique opératoire du chirurgien pouvaient influencer la conduite d'un remaniement de son geste opératoire.²²⁶ Cet exemple illustre bien la nécessité de prendre en considération un tel effet d'apprentissage lors de la conduite d'études comparatives sur l'efficacité d'une nouvelle technique chirurgicale.^{227,78,228}

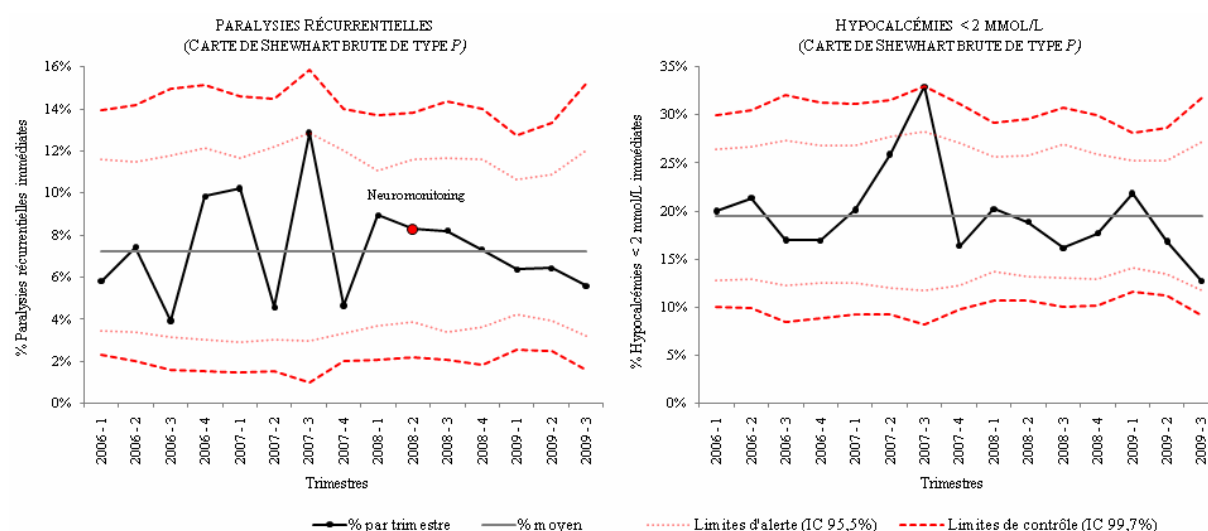


Figure 14. Cartes de contrôle représentant l'évolution de la proportion trimestrielle de complications postopératoires immédiates liées à la chirurgie thyroïdienne

Apports méthodologiques liés à l'élaboration des cartes de contrôle

L'usage combiné des différents types de cartes de contrôle a rendu possible une analyse plus complète des résultats de la chirurgie thyroïdienne (table XVI). L'emploi de la carte de Shewhart était adapté au suivi permanent de la performance d'ensemble de l'équipe chirurgicale, du fait de sa simplicité de construction et d'interprétation. Au regard des bénéfices apportés par l'utilisation de la carte de contrôle brute en routine, le développement d'une carte adaptée au suivi des complications de la chirurgie thyroïdienne s'est avéré certes élégant mais peu convaincant. Ce travail a montré l'intérêt modéré de la carte ajustée pour

détecter des causes spéciales de variation. Compte tenu des données dont nous disposions, un tel résultat n'est peut-être pas surprenant. Il traduit à la fois la connaissance limitée et une prise en compte insuffisante des facteurs de variations influençant le résultat de cette chirurgie. Un approfondissement des modèles d'ajustement est envisageable pour démontrer l'intérêt d'utilisation de la carte de Shewhart ajustée en pratique courante. En gommant des facteurs de variations non liés à la qualité du soin, elle doit conduire à une plus grande fiabilité de l'outil, c'est-à-dire à améliorer sa sensibilité pour la détection de causes spéciales et sa spécificité à ne pas détecter des variations insignifiantes de l'indicateur.^{113,114}

La carte CUSUM était réservée à l'interprétation occasionnelle de la performance de chacun des chirurgiens. Elle s'est avérée utile pour analyser les variations de faible amplitude des indicateurs sur un nombre limité de patients opérés.¹¹⁵ Son emploi était particulièrement approprié lorsque nous avons étudié l'apprentissage par les chirurgiens du neuromonitoring. Pour chacun d'entre eux, disposer d'un retour individualisé sur sa propre performance facilitait la remise en cause de sa manière d'opérer et ainsi la possibilité de toujours continuer à progresser. Toutefois, la construction comme l'interprétation de cette carte demeuraient complexes et ont nécessité un accompagnement soutenu du méthodologiste pour que les chirurgiens se la réapproprient.

Table XVI. Principes d'utilisation des cartes de Shewhart et CUSUM

	Carte de Shewhart	Carte CUSUM
Évaluation	Collective	Individuelle
Construction	Facile	Difficile
Interprétation	Intuitive	Complexe
Variations	Grande amplitude	Faible amplitude
Processus	Instable	Stable
Données	Échantillons de cas groupés	Série de cas
Ajustement	Recommandé	Nécessaire

6.1.2. Limites

Généralisation des résultats

Ce travail a été conduit dans un seul centre, ce qui limite forcément la représentativité de nos résultats à la pratique des trois chirurgiens endocriniens expérimentés dont nous avons suivi l'activité. De plus, nous nous sommes intéressés à l'évaluation d'une chirurgie particulière : la chirurgie de la thyroïde. Cette chirurgie était caractérisable par le volume élevé d'interventions réalisées, une technique opératoire relativement standardisée, l'homogénéité des populations de patients opérés et la fréquence élevée des complications immédiates surveillées.¹⁷⁵ Elle se rapprochait donc davantage du milieu d'application initial des cartes de contrôle, l'industrie manufacturière, que d'autres types de chirurgie. Bien que nos découvertes servent la chirurgie thyroïdienne, la prudence s'impose lorsque l'on souhaite les extrapoler à d'autres prises en charge. Même si l'on peut espérer obtenir des résultats similaires en appliquant les cartes à des soins fréquents et suffisamment normalisés, leur reproductibilité reste à démontrer dans diverses spécialités chirurgicales ou médicales. L'intérêt de la carte de contrôle doit être étudié lorsqu'elle est appliquée au sein d'équipes de chirurgiens moins expérimentés, ainsi que pour le suivi de complications rares ou consécutives à un geste chirurgical exceptionnel, difficilement reproductible. Un volume minimum d'interventions est probablement nécessaire pour garantir une précision suffisante de la mesure des indicateurs, ce qui contraint l'utilisation de cartes de Shewhart lorsque l'on souhaite suivre la performance individuelle du chirurgien.

Compte tenu de notre schéma d'étude, nous n'avons pas pu objectiver l'efficacité d'une utilisation routinière des cartes de contrôle pour réduire la survenue de complications opératoires. En l'absence de groupe témoin contemporain à la conduite de cette expérience, notre analyse dynamique de l'évolution des résultats chirurgicaux au cours du temps ne permettait pas le contrôle de facteurs de confusion potentiels ayant pu influencer les tendances observées sur les indicateurs.²²⁹ Néanmoins, les modalités de détection des complications et leur recueil sont restés inchangés et conformes au protocole d'évaluation pendant toute la durée de l'étude. Nous nous sommes également efforcés de tenir compte des variations de recrutement du service par un ajustement des indicateurs sur les caractéristiques des patients opérés. Bien que non répertoriés dans le journal de bord ou difficiles à contrôler, d'autres facteurs propres au chirurgien, à l'équipe chirurgicale ou à l'organisation du bloc opératoire ont pu agir sur les résultats obtenus.

Le caractère monocentrique et non expérimental de notre étude limite la transposition de ses résultats à d'autres sites et d'autres chirurgies. En optant pour un design observationnel, nous souhaitons mieux comprendre la complexité des facteurs pouvant conditionner le résultat d'une chirurgie thyroïdienne en pratique courante. Ainsi, nous espérons révéler des hypothèses de recherche à vérifier secondairement dans le cadre de protocoles d'étude plus larges.²³⁰

Qualité des données utilisées

Les données utiles à ce travail provenaient d'une exploitation des bases du Programme de Médicalisation des Systèmes d'Information (PMSI). Souvent débattu, l'usage de ces données à des fins de recherche impose de bien connaître leurs règles de codage et les limites de leur recueil.²³¹ Parmi les données habituellement renseignées dans le PMSI, nous n'avons utilisé que celles dont nous pensions obtenir une qualité et une exhaustivité de codage satisfaisantes. Les informations que nous jugions essentielles pour mener une évaluation rigoureuse, incluant la survenue des complications opératoires, provenaient d'un recueil complémentaire instauré au démarrage du suivi prospectif.

Le nombre de variables prévues dans le protocole d'évaluation était volontairement restreint afin de ne pas alourdir un recueil de données s'ajoutant aux tâches quotidiennes des chirurgiens. Plutôt que le sexe et l'âge du patient opéré, la nature approximative de la pathologie thyroïdienne ou de l'acte réalisé, d'autres variables auraient pu optimiser la capacité de nos modèles d'ajustement à capter une part plus importante de la variabilité des résultats de la chirurgie. Bien que leur mesure soit délicate, on peut citer certains facteurs liés aux caractéristiques de la chirurgie (présence d'un goitre large ou plongeant, d'un cancer infiltrant), au profil du chirurgien (technique opératoire, état de forme physique et mentale), ou à l'organisation du bloc opératoire (équipements disponibles, coordination d'équipe, planning opératoire ou charge de travail).⁸⁰ L'identité du chirurgien n'était pas recueillie avant la mise en place du neuromonitoring, ce qui explique qu'elle ne figurait pas dans le set des variables d'ajustement.

La mesure des résultats de la chirurgie thyroïdienne reposait sur la survenue de complications recherchées immédiatement après l'intervention. La fréquence des complications opératoires était particulièrement élevée compte tenu de la sensibilité de notre méthode de dépistage et de la présence de nombreuses complications transitoires se résolvant

spontanément au cours du temps.¹⁸⁴ Idéalement, la sécurité d'une chirurgie doit être évaluée en fonction des complications permanentes qu'elle peut occasionner et qui sont susceptibles d'affecter directement la santé du patient. Après une thyroïdectomie, le caractère définitif d'une paralysie récurrentielle ou d'une hypoparathyroïdie secondaire ne peut être affirmé avant un délai minimum de 6 mois. Le retentissement de ces complications sur la qualité de vie du patient est difficilement évaluable et peut dépendre fortement de son ressenti ou de ses activités quotidiennes. Nous n'avons pas pu mettre en œuvre un tel suivi des complications à long terme pour des raisons de faisabilité évidentes et parce qu'il n'était pas pertinent de retarder la restitution des cartes, ce qui aurait nuit à la réactivité de notre démarche. Ne disposant pas d'un examen capable de prédire, au décours immédiat de l'intervention, le caractère définitif de la complication, la conduite de travaux de validation s'impose pour déterminer si les variations observées sur les indicateurs de complications immédiates concordent bien avec celles attendues à distance de l'acte.

Notre protocole d'évaluation s'appuyait sur la réalisation systématique d'une laryngoscopie directe qui constitue l'examen de référence pour la détection des paralysies récurrentielles postopératoires. En l'absence de *gold standard* établi pour dépister les hypoparathyroïdies secondaires, nous avons opté pour un dosage de la calcémie totale le jour suivant l'intervention, répété en cas de résultat anormal lors du premier dosage (calcémie < 2 mmol/L).^{192,193} Bien que rarement observée, une hypocalcémie peut survenir de manière décalée après l'intervention, se traduisant par une calcémie normale le lendemain de l'acte chirurgical mais anormale le surlendemain. Un tel phénomène a pu être omis par le protocole d'évaluation. La probabilité de détecter une hypocalcémie dépend du seuil de détection fixé et de la fréquence des mesures effectuées dans les 48 heures suivant l'intervention. Par conséquent, un nombre minimum de dosages, voire leur séquençage régulier dans le temps, permettrait d'améliorer la fiabilité de l'indicateur. Des contraintes logistiques peuvent néanmoins limiter la faisabilité d'un tel protocole, puisque le patient est susceptible de quitter l'établissement précocement en l'absence de complications détectées le lendemain de l'opération. Enfin, bien que la mesure de la calcémie corrigée n'ait pas été retenue à des fins de simplification, il conviendrait d'examiner plus en détail les avantages et difficultés inhérentes à la réalisation de cet examen en période postopératoire.

6.2. Perspectives

6.2.1. Application des cartes de contrôle aux hôpitaux français

Développer un indicateur pour mesurer le résultat d'une chirurgie est un travail périlleux compte tenu de sa difficulté à refléter la qualité des prises en charge réalisées et des enjeux liés à son utilisation.⁴⁶ Il doit être mené préférentiellement pour une intervention fréquente dont les complications sont facilement quantifiables. Tout d'abord, le choix de l'indicateur comme celui des variables d'ajustement doit être réfléchi³⁶ et consensuel afin qu'il soit accepté par les professionnels concernés par l'évaluation. Idéalement, l'élaboration d'un score de risque préopératoire spécifique est utile pour pondérer l'indicateur en fonction de la diversité des situations rencontrées (caractéristiques du patient, sévérité de la pathologie traitée, complexité de l'acte réalisé, etc.).^{232,233,72} Une fois l'indicateur construit, il convient d'en valider la mesure ainsi que la variation de cette mesure, ce qui demande du temps et des moyens conséquents.²³⁴ Ce n'est qu'à l'issue de ces étapes successives que son utilisation peut être envisagée en routine.

Les hôpitaux constituent un terrain privilégié lorsque l'on souhaite homogénéiser la qualité et la sécurité des soins dispensés à la population d'un pays. Les prises en charge qu'ils délivrent sont caractérisables et présentent des volumes suffisants pour espérer une mesure valide. De plus, les hôpitaux disposent de leurs propres systèmes d'information et de structures de gouvernance locales utiles pour mettre en œuvre un programme d'évaluation des soins.²³⁵ A l'image du modèle anglais, deux approches complémentaires reposant sur l'emploi de cartes de contrôle sont possibles pour améliorer leurs résultats. La première consiste à comparer la performance de plusieurs hôpitaux en étudiant la répartition d'un indicateur de résultat à l'aide d'une carte transversale (*funnel plot*).²³⁶ Plutôt que de classer imparfaitement ces hôpitaux, une telle présentation des résultats discrimine ceux qui s'écartent significativement d'une performance moyenne jugée acceptable et donc à haut risque pour le patient. La conduite d'investigations sur le terrain permet alors d'identifier les causes de la mauvaise performance de certains établissements pour les aider à améliorer la sécurité de leurs prestations en apprenant de leurs erreurs.⁴⁷ La seconde approche repose sur la mise à disposition de cartes longitudinales (cartes de Shewhart ou CUSUM)²³⁷ pour suivre les variations d'un indicateur au cours du temps. Ainsi, chaque hôpital est amené à interpréter ses résultats de manière autonome, ce qui doit l'inciter à en réduire la variabilité puis à les

améliorer. S'inspirant de la pensée de Shewhart, l'usage combiné des cartes de contrôle transversales et longitudinales^{106,238} autorise donc une surveillance nationale de la sécurité des soins chirurgicaux et guide la conduite d'actions d'amélioration au sein même des hôpitaux.

La mise en place d'un tel dispositif est faisable en France. Les données du PMSI représentent un matériel de choix. Leur collecte est généralisée à l'ensemble des hôpitaux et leur qualité de plus en plus contrôlée. Les pays de l'Organisation de Coopération et de Développement Économiques (OCDE) comparent déjà la sécurité de leurs systèmes hospitaliers à l'aide d'indicateurs extraits directement de bases médico-administratives (*Patient Safety Indicators*).⁴⁴ Parmi ces indicateurs figurent notamment les complications d'anesthésie, la septicémie, l'embolie pulmonaire ou la thrombose veineuse profonde postopératoires.²³⁹ De nombreux travaux, auxquels nous participons activement,²⁴⁰ sont actuellement menés à travers le monde pour en valider la mesure.^{234,241,242}

6.2.2. Le chirurgien, l'artisan et le pilote

Les facteurs qui concourent au résultat d'une chirurgie sont encore mal connus aujourd'hui et peu étudiés. L'essentiel des efforts se concentre sur les innovations technologiques influençant la technique chirurgicale ou les modalités d'anesthésie permettant de mieux tenir compte du profil des patients. Pourtant, il paraît difficilement envisageable que la robotique, la chirurgie mini-invasive ou les tentatives visant à tayloriser la prise en charge au bloc, remplacent un jour l'action de l'homme ou parviennent à en gommer complètement les erreurs. La complexité des éléments conditionnant la performance du chirurgien ou de l'équipe chirurgicale explique peut-être ce manque d'intérêt pour le facteur humain. Son impact sur la sécurité du patient est néanmoins déterminant.⁴⁵

Lorsque l'on souhaite simplifier le problème en établissant une typologie de la performance chirurgicale, deux composantes paraissent incontournables : la nature du geste opératoire effectué, routinier ou exceptionnel, et l'implication variable des différents métiers présents en salle d'opération (table XVII). Les méthodes mises en œuvre pour mesurer et améliorer efficacement la sécurité du patient opéré diffèrent probablement selon ces types de chirurgie. Le suivi d'indicateurs par cartes de contrôle semble particulièrement adapté aux interventions de routine, pour lesquelles la performance chirurgicale est plus facilement quantifiable et contrôlable que pour des interventions exceptionnelles.⁹⁶ Les efforts à fournir pour réduire la survenue des complications opératoires devront cibler plus précisément le chirurgien ou bien l'ensemble de l'équipe au bloc opératoire, selon leur contribution variable au résultat d'une chirurgie. Cela implique de mieux comprendre les facteurs pouvant expliquer la variabilité de leurs performances respectives.

Table XVII. Classement des chirurgies selon leur fréquence et le niveau d'implication de l'équipe de bloc opératoire dans leur réalisation

	Performance individuelle du chirurgien	Performance collective de l'équipe chirurgicale
Intervention de routine	Appendicectomie	Pontage aorto-coronarien
Intervention exceptionnelle	Exérèse de tumeur parotidienne	Surrénalectomie

Afin d'étudier l'évolution de la performance individuelle du chirurgien au cours de sa carrière, il convient d'élargir notre angle d'approche à d'autres domaines. Dans tout métier, le passage d'une connaissance générale théorique à l'acquisition d'une compétence motrice est progressif. Au début, le novice doit comprendre et intellectualiser la tâche à accomplir, puis en appliquer rigoureusement chacune des étapes successives. Avec la pratique, l'apprenti parvient à intégrer le geste ce qui lui permet de gagner en efficacité et en rapidité jusqu'à ce qu'il devienne expert, c'est-à-dire qu'il automatise la technique et soit capable de la reproduire de manière fluide et intuitive.²⁴³ L'expertise représente le degré le plus élevé d'acquisition d'une compétence et le résultat final d'une amélioration progressive de la performance du professionnel, qui est supérieure à la moyenne. Depuis la Grèce Antique, cette expertise est traditionnellement approchée en fonction de l'expérience accumulée par le professionnel, sous-entendant que le savoir et le savoir-faire augmentent nécessairement avec l'âge.²⁴⁴ Toutefois, des recherches ont récemment suggéré que performance et expérience du professionnel n'étaient pas nécessairement corrélées, conduisant à une sérieuse remise en question de cet adage.²⁴⁵ On peut s'attendre à ce que la réalisation de tâches basiques ne s'accompagne pas systématiquement d'une amélioration de la performance du médecin calquée sur le nombre d'années d'exercice suivant sa formation initiale, sa compétence pouvant même décliner après l'obtention du diplôme.²⁴⁶

Selon Anders Ericsson, il existe un ordonnancement dans la difficulté des tâches à maîtriser avant d'atteindre un niveau d'expertise reconnu dans de nombreux domaines tels les échecs ou la musique. Même les individus les plus talentueux doivent s'impliquer intensément pour espérer devenir experts, ce qui peut prendre plusieurs décennies. Ainsi, l'âge auquel ils atteignent leur pic de performance se situerait typiquement entre 30 et 50 ans, la plupart tentant ensuite d'en maintenir le niveau stable durant le reste de leur carrière (figure 15). Plutôt que de reposer uniquement sur l'expérience du professionnel et l'antériorité de sa pratique, des gains de performance significatifs sont obtenus quand il est motivé pour s'améliorer sur une tâche spécifique dont il peut répéter et affiner la réalisation. Cela demande une grande concentration et implique d'entrer dans une dynamique de résolution de problèmes, plus facilement envisageable lorsque le professionnel dispose d'un retour en temps réel sur son activité. La chirurgie offre de telles possibilités de progression, puisque l'opérateur observe immédiatement la conséquence de ses gestes et de ses éventuelles erreurs.²⁴⁷ Les experts y sont habituellement désignés comme des chirurgiens expérimentés présentant de meilleurs résultats que leurs collègues non experts. Cette définition est soutenue par les données de la littérature sur le lien entre volume et résultat de la chirurgie bien qu'il

puisse exister des différences entre praticiens à haut volume d'activité, suggérant que tous ne peuvent devenir experts.²²⁶ Les chirurgiens experts se distinguent alors par leur capacité à savoir quoi faire et comment le faire, à projeter leur raisonnement sur la réalisation du geste opératoire tout en sachant corriger leurs erreurs et reproduire leur performance.²⁴⁸

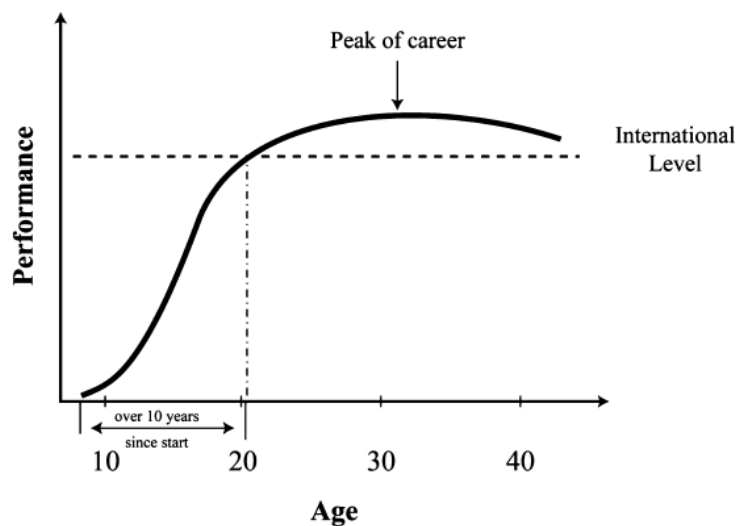


Figure 15.

Amélioration progressive de la performance d'un joueur d'échec en fonction de son âge avant d'atteindre un niveau d'expertise internationale

[Extrait de K.A. Ericsson]²⁴⁹

Il est probable que la cinétique caractérisant l'atteinte et le maintien d'un niveau d'expertise en chirurgie diffère sensiblement de ce qui est observé dans les domaines artistique ou sportif. Si le résultat d'une chirurgie dépend avant tout de la compétence du chirurgien plutôt que de son étroite collaboration avec les autres membres du bloc opératoire, s'il opère seul avec l'unique apport 'de ses mains et de sa tête', son travail s'apparentera à celui d'un artisan. Plus précisément, on pourra le rapprocher du menuisier pour ce qui relève d'un geste opératoire ordinaire, c'est-à-dire une chirurgie fréquente, relativement simple et bien standardisée (figure 16, intervention de routine). Dans ce cas, on peut s'attendre à ce qu'il atteigne un pic de performance en milieu de carrière, une fois sa formation initiale terminée et la technique chirurgicale bien maîtrisée. Puis, en l'absence d'entraînement adapté, de remise en question régulière, on peut s'attendre à voir sa performance décliner en partie due à une lassitude mentale liée à la répétition du geste. L'inscription à des modules d'enseignements pratiques ou l'introduction d'innovations technologiques, tels la cœlioscopie ou le neuromonitoring, peut alors faciliter un réapprentissage ou une réinvention du geste opératoire (figure 16, réapprentissage du geste). Partant du principe que la plasticité du geste opératoire est inversement proportionnelle à son degré de systématisation, son remodelage sera d'autant plus difficile que la technique habituelle du chirurgien est déjà bien rôdée, acquise et reproduite depuis longtemps. A la lueur de nos travaux, ce réapprentissage peut

prendre du temps, n'est pas nécessairement progressif mais peut nécessiter plusieurs paliers et une dégradation temporaire de la performance du chirurgien avant que la nouvelle technique soit maîtrisée et que ses résultats s'améliorent.

Si l'on s'intéresse à un geste opératoire rarement réalisé, difficilement normalisable et plus proche d'un travail d'ébénisterie, le principal gain de performance à escompter proviendra de l'expérience du chirurgien. Sa capacité à poser prudemment l'indication chirurgicale, à planifier les temps successifs de l'intervention par une représentation mentale anticipant d'éventuels écueils, primera sur l'indécision du jeune chirurgien. Une amélioration progressive des résultats du chirurgien sur ce type d'actes complexes est donc probable tout au long de sa carrière, se bonifiant au fur et à mesure de la diversité des situations rencontrées (figure 16, intervention exceptionnelle). Bien sûr, il ne s'agit encore que d'hypothèses, qui nécessitent d'être étayées de preuves solides avant d'envisager d'éventuelles applications pratiques. Ces applications concernent aussi bien les modalités de formation des chirurgiens au cours de leur carrière, que la répartition des interventions de routine ou exceptionnelles entre opérateurs d'une même équipe chirurgicale. La conduite d'études visant à mesurer l'évolution de la performance du chirurgien tout au long de sa vie professionnelle, en fonction du type de chirurgie réalisée, paraît justifiée malgré une durée de suivi prolongée en limitant la faisabilité.²⁴⁵

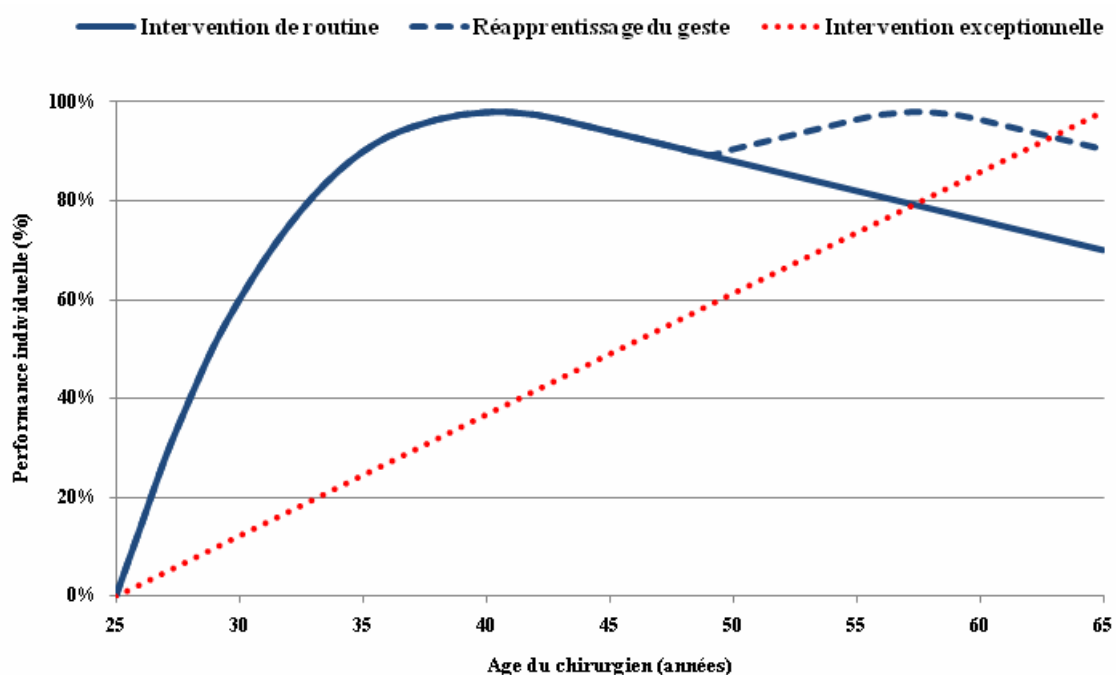


Figure 16. Évolution de la performance individuelle du chirurgien au cours de sa carrière en fonction du type d'intervention réalisée [inspiré de l'approche de K.A. Ericsson]²⁴⁹

Une telle représentation de l'art chirurgical est évidemment réductrice et doit intégrer, en plus des caractéristiques du patient opéré, le talent propre du chirurgien, la conjonction de ses capacités manuelles et intellectuelles, son background académique et professionnel, mais aussi sa personnalité qui conditionnera ses prises de décision et ses réactions face à l'imprévu. Dans le cadre d'une intervention de routine, un chirurgien besogneux, s'efforçant d'appliquer scrupuleusement la conduite à tenir opératoire sera sans doute plus performant qu'un chirurgien génial mais imprévisible. A l'inverse, lors d'une intervention exceptionnelle, un chirurgien imaginatif, adaptable et capable de gérer son stress dans une situation inhabituelle, obtiendra de meilleurs résultats qu'un chirurgien inexpérimenté à la gestuelle stéréotypée. A cela peut s'ajouter l'état de forme de l'opérateur un jour donné. Une fatigue occasionnelle ou plus insidieuse, suite à une période de suractivité prolongée ou à la survenue d'un accident iatrogène au bloc opératoire, peut se répercuter directement sur son mental.¹⁵ Au-delà d'entraîner une baisse de sa vigilance, elle représente une source de stress qui peut l'amener à douter de la sûreté de son geste ou de son jugement clinique. Comme dans tout métier, la prise en compte du facteur humain est essentielle au développement des compétences du chirurgien. Le talent et l'expérience ne suffisent pas pour garantir la sécurité d'une chirurgie si la motivation du chirurgien et sa capacité à progresser sont défaillantes, car c'est de la combinaison de tous ces facteurs que dépend sa performance individuelle.

Bien que la compétence technique du chirurgien demeure un pré-requis essentiel pour une chirurgie sûre, ses aptitudes à communiquer et à travailler en équipe sont également déterminantes lorsque le résultat d'une chirurgie dépend de la performance d'ensemble de l'équipe de bloc opératoire.²⁵⁰ Dans ce cas, l'activité du chirurgien peut s'apparenter à celle d'un pilote de ligne. Elle doit intégrer de nombreux paramètres extérieurs, qui relèvent à la fois de l'organisation logistique au sein du bloc opératoire (planning opératoire cohérent, installation appropriée du patient, disponibilité des équipements, etc.), du management des professionnels participant à l'intervention (anesthésiste, personnel infirmier, etc.) et du degré d'urgence de la prise en charge. Ses capacités d'observation et de leadership seront donc primordiales pour parvenir à un travail d'équipe efficace.²⁵¹ L'augmentation du nombre d'intervenants au bloc tend à accroître le risque de complications chirurgicales du fait de la multiplication d'erreurs susceptibles d'intervenir au niveau des interfaces entre les métiers. Paradoxalement, elle représente autant de gardes-fous pour détecter à temps une éventuelle défaillance de l'un des participants. C'est pourquoi une bonne connaissance de la nature des rapports entre les personnes et un encadrement de leurs modes de communication peuvent également améliorer la sécurité du patient opéré.

6.2.3. Futurs développements liés à la sécurité du patient en chirurgie

L'OMS a fixé en 2009 des axes de recherche prioritaires sur la sécurité du patient en chirurgie. Le manque de communication et de coordination, des défauts d'organisation latents et une culture 'sécurité' insuffisante au sein des blocs opératoires arrivaient en tête d'une liste de problèmes à investiguer dans les pays développés.²⁵² Les effets de la pression au travail, du stress et de la fatigue ressentis par les professionnels de santé y figuraient également.⁴⁵

Volume d'activité et fatigue du chirurgien

De nombreux travaux ont tenté de déterminer un seuil minimum d'activité à atteindre pour garantir des soins chirurgicaux avec un niveau de sécurité suffisant. Ces travaux ont montré qu'un volume important d'activité pouvait être associé à une moindre mortalité, selon le type d'intervention considéré.^{28,29,30,31} Bien que cette association soit explicable par un effet d'apprentissage (*practice makes perfect*) ou par une orientation préférentielle des patients vers les établissements ayant de bons résultats (*selective referral*),²⁵³ d'autres facteurs peuvent également intervenir (caractéristiques des patients, facteurs organisationnels ou liés aux professionnels).²⁵⁴ De plus, l'impact du volume sur le résultat des soins semble s'atténuer au fur et à mesure que l'activité augmente²⁵⁵ et, comme cela a pu être montré en deçà d'un certain seuil, il existe probablement un plafond d'activité au-delà duquel surviendrait une altération de la sécurité des soins. Cette préoccupation est renforcée par la généralisation dans la plupart des pays industrialisés d'un mode de financement des hôpitaux à la pathologie,²⁵⁶ qui vise à augmenter l'activité hospitalière.²⁵⁷ Si la réduction des durées de séjour peut s'accompagner d'une exposition plus courte du patient au risque d'accident iatrogène, une augmentation excessive de l'activité chirurgicale peut également entraîner l'instabilité du processus de prise en charge et un surmenage des professionnels de santé, se traduisant par une augmentation des complications liées à l'intervention ou à l'épisode hospitalier.⁴⁵

La place du médecin à l'hôpital est centrale. Son bien-être et sa satisfaction au travail sont essentiels pour garantir des soins de qualité et limiter les risques d'erreurs humaines.²⁵⁸ Des conditions d'exercice de plus en plus stressantes et un manque d'autonomie croissant dans ses prises de décisions, liés au contrôle accru des autorités de santé, des employeurs comme des patients, sont susceptibles d'altérer sa performance.²⁵⁹ Une longue période de surmenage peut ainsi mener au *burnout*, à la dépression ou à une détresse psychologique générale et se répercuter directement sur les soins délivrés aux patients. Le *burnout* est un

syndrome qui comprend trois phases successives : d'abord un épuisement émotionnel dû à un stress intense et permanent avec une perte d'enthousiasme pour les tâches courantes, puis une dépersonnalisation c'est-à-dire une distanciation par rapport au patient, enfin un reniement lié à la perception d'un faible accomplissement de soi au travail.²⁶⁰ Ses causes sont multiples et regroupent la surcharge de travail, une multiplication des contraintes administratives, des impératifs de productivité, une médecine judiciarisée ou l'impression de ne plus être reconnu à sa juste valeur. Sa conséquence habituelle est la baisse d'efficacité du professionnel dans le cadre de ses activités quotidiennes, ce qui peut altérer sa relation au patient et la qualité de ses prestations.²⁶¹ Bien que de telles manifestations puissent toucher le professionnel à tous les stades de sa carrière, les internes sont particulièrement concernés. Le *burnout* est fréquent dans cette population et peut entraîner des erreurs de prescriptions ou de prises en charge.^{262,263,264} De nombreux travaux ont mis en évidence que leur manque de sommeil ou des plages de travail continu trop étendues constituaient un danger à la fois pour le patient^{265,266} et pour eux-mêmes.²⁶⁷ Ces constatations ont récemment incité l'Institut de Médecine Américain à limiter la durée maximale de travail des internes à 80 heures par semaine, sans dépasser un plafond fixé à 30 heures d'exercice continu incluant des plages de repos.²⁶⁸ De telles mesures se révèlent toutefois difficilement applicables et demeurent largement débattues outre-Atlantique.^{269,270}

Les tâches réalisées par un chirurgien varient selon son lieu d'exercice. Alors qu'un chirurgien exerçant dans une structure académique doit répondre à un éventail d'activités cliniques, administratives, d'enseignement et de recherche, un chirurgien libéral est davantage disponible pour concentrer son attention sur la pratique des soins.²⁷¹ Selon une étude récente, plus d'un tiers des chirurgiens américains se déclarent être en *burnout* dont la survenue serait associée au jeune âge de l'opérateur, au fait d'avoir des enfants, à la spécialité exercée, au nombre d'appels reçus la nuit ou d'heures de travail hebdomadaires.²⁷² Les chirurgiens rapportant une erreur de prise en charge récente seraient ainsi davantage sujets au *burnout* ou à la dépression.²⁷³ La fatigue du chirurgien conditionne sa performance. Elle dépend principalement de sa charge de travail, d'une éventuelle privation de sommeil ou du rythme circadien. Elle peut se répercuter directement sur la dextérité et l'exactitude de son geste.²⁷⁴ Le risque de complications opératoires augmenterait après une nuit de garde durant laquelle le chirurgien a dormi moins de 6 heures,²⁷⁵ en cas d'intervention réalisée en début d'après-midi²⁷⁶ ou en fin de semaine,²⁷⁷ ou bien après 24 heures de travail continu.²⁷⁸ Cette fatigue peut être amplifiée au sein d'un environnement stressant qui altèrera la vigilance du

chirurgien, affectera sa réalisation de tâches complexes et retentira sur ses relations avec les autres professionnels du bloc opératoire.²⁷⁹

Un planning opératoire inadapté ou une suractivité chirurgicale prolongée représentent un risque de fatigue excessive, de lassitude ou de *burnout* pour le chirurgien. Certes, un volume d'interventions minimum est nécessaire pour qu'il obtienne des résultats acceptables et un équilibre doit sans doute être trouvé entre sous-activité et suractivité chirurgicale. Des recherches demeurent nécessaires pour déterminer s'il existe une fenêtre d'activité optimale, incluant des seuils à ne pas dépasser, pour garantir la sécurité des patients opérés. Ces travaux ne doivent pas se limiter à une vision macroscopique se basant uniquement sur l'étude des volumes d'activité annuels par hôpital, mais aussi considérer la gestion quotidienne des programmes opératoires et les facteurs influençant la performance individuelle du chirurgien. L'ensemble de ces réflexions est en cours d'étude, dans le cadre d'un projet de recherche que nous menons au sein de cinq centres français présentant une forte activité de chirurgie thyroïdienne (étude observationnelle CATHY).

Apports de l'aéronautique à la coordination du travail des équipes au bloc

Les défauts de communication entre professionnels du bloc opératoire sont classiques²⁸⁰ et peuvent occasionner des erreurs de prise en charge à l'origine de complications pour le patient.^{87,281} Plutôt que de rester focalisé sur la technique chirurgicale ou les innovations technologiques, des marges de progression importantes sont attendues d'une meilleure prise en compte du facteur humain et de l'organisation des prises en charge au bloc. La performance collective d'une équipe chirurgicale dépend en grande partie de sa bonne coordination. Cela requiert l'aptitude des professionnels au travail en équipe et à la communication interpersonnelle, des compétences auxquelles ils ne sont habituellement pas formés ou pour lesquelles leur intérêt paraît limité.²⁸²

L'adaptation au domaine chirurgical des méthodes développées dans l'aviation civile laisse entrevoir des gains importants pour apporter aux équipes une culture de sécurité du patient et ainsi éviter la survenue d'accidents au bloc opératoire.²⁸³ Deux modèles de sécurité en vol ont été développés dans l'aéronautique : le modèle normatif et le modèle adaptatif.²⁸⁴ Le modèle normatif propose de réduire les erreurs humaines par une standardisation des tâches à accomplir. Le modèle adaptatif concerne des activités se prêtant mal à la

standardisation du fait de nombreuses situations d'imprévu qui requièrent une excellente cohésion et une forte communication au sein des équipes. Ces deux modèles peuvent être utiles en chirurgie de manière combinée.

Les premières listes de contrôle (ou *check-lists*) développées dans l'aéronautique s'inscrivent dans le cadre du modèle normatif et ont démontré leur efficacité pour réduire les risques d'accidents.²⁸⁵ Bien que récemment introduites et d'appropriation relativement lente dans le secteur de la santé,^{281,286} l'utilisation de *check-lists* ou de briefings préopératoires systématiques entre les professionnels du bloc^{287,288} est prometteuse.²⁸⁹ D'abord conçu pour décharger les équipes d'efforts de mémorisation inutiles et ainsi éviter les erreurs d'omission,²⁹⁰ l'usage de *check-lists* permet une standardisation des procédures, rappelant des points-clefs à vérifier systématiquement. Toutefois, il ne garantit pas une haute sécurité lorsque la standardisation des tâches s'avère irréalisable ou inappropriée (situations d'urgence, prises en charge complexes). Certes, les *check-lists* permettent de sécuriser l'exécution des soins en vérifiant que ce qui a été décidé est fait, mais elles ne permettent pas de sécuriser la décision de soins qui nécessite d'autres approches répondant au modèle adaptatif.

Dans le but de mieux maîtriser les facteurs humains impliqués dans la survenue d'erreurs de pilotage, l'aviation civile a développé depuis 1979 le concept du *Crew Resource Management* (CRM).^{291,292} Le CRM repose sur un entraînement formel des équipes à l'échange d'informations. Sa finalité est d'aboutir au partage d'un modèle mental commun du vol par les différents membres de l'équipage pour prévenir et mieux gérer la survenue de problèmes de sécurité. Le CRM inclut des briefings durant lesquels la personne responsable passe en revue les tâches à réaliser et souligne les écueils potentiels à éviter. Chacun des membres de l'équipe est également amené à se manifester s'il détecte une menace potentielle à la sécurité du vol.²⁹³ La perception de la situation présente, la compréhension de cette situation et la projection des événements à venir sont essentielles.²⁹⁴ Le CRM s'inscrit donc dans le cadre du modèle adaptatif dès lors qu'il facilite la détection précoce d'un imprévu et contribue à y répondre avec flexibilité. En pratique, le CRM consiste en une formation théorique, lors de laquelle des grands principes de *team management* sont expliqués aux professionnels. L'apprentissage des connaissances est ensuite testé en simulateur de vol. Habituellement, l'équipage doit gérer une mission scénarisée au cours de laquelle vont survenir certains imprévus, ce qui permet de développer des habitudes et de mettre en pratique des règles. À l'issue de cette simulation, un débriefing est mené au moyen d'enregistrements

vidéo, pour faire prendre conscience à l'équipage de ses qualités et faiblesses, à la fois dans les tâches routinières et les situations d'urgence.²⁹⁵

Au-delà d'une simple aide technique, la *check-list* représente un outil facilitant la collaboration entre disciplines.²⁹⁶ En incitant à la communication, son emploi ouvre des perspectives importantes pour parvenir à mieux structurer les échanges d'informations entre personnes au bloc opératoire.²⁹⁷ L'adjonction de méthodes d'entraînement au CRM représente un moyen d'accroître l'adhésion des professionnels à la *check-list* et d'améliorer leur travail en équipe en vue de réduire le risque d'erreurs humaines.^{298,299,300,301} Certains hôpitaux les ont déjà implémentées³⁰² et l'*American College of Surgeons* a récemment recommandé l'entraînement des chirurgiens au CRM.³⁰³ Bien que de nombreux travaux aient démontré la capacité du CRM à changer favorablement les comportements des professionnels de bloc opératoire,³⁰⁴ peu ont évalué son efficacité sur le résultat des soins. On peut s'attendre à ce que son effet soit plus important lors d'opérations complexes (transplantation d'organes) que lors d'interventions où l'opérateur est peu tributaire d'autrui (cholécystectomie). Avant d'envisager la généralisation du CRM à l'ensemble des équipes de bloc opératoire, la conduite d'études à haut niveau de preuve s'impose donc pour déterminer sa capacité à réduire la survenue d'événements indésirables liés aux soins chirurgicaux et pour en préciser les méthodes d'entraînement.³⁰⁵ Nous tentons actuellement d'obtenir le financement nécessaire à la conduite d'une étude d'intervention multicentrique qui vise à mesurer l'impact d'un entraînement des équipes de bloc au CRM.

7. Conclusion

Le résultat d'une chirurgie thyroïdienne dépend de la combinaison d'une multitude de facteurs. Bien que la plupart demeurent encore inconnus, la carte de contrôle participe à leur découverte. Près d'un siècle après sa création dans l'industrie manufacturière, la carte de Shewhart contribue désormais à améliorer la sécurité des prises en charge chirurgicales. Il s'agit d'un outil simple d'utilisation et peu onéreux à mettre en œuvre qui, en guidant l'interprétation d'un indicateur, génère de la connaissance sur ses causes de variations afin de mieux les maîtriser. Cependant, l'application des méthodes de contrôle qualité statistiques aux soins ne doit pas empiéter sur l'autonomie des professionnels et requiert une certaine prudence. La réalisation d'une chirurgie demeure une tâche complexe, hautement variable. Les éléments qui concourent naturellement à son résultat doivent être considérés méthodiquement pour déterminer ce qui constitue la preuve d'un changement réel de la prise en charge. L'introduction dans les meilleures conditions d'indicateurs pertinents et valides ne suffit pas pour réduire la survenue de complications chirurgicales. Seule une analyse rigoureuse de leur variabilité peut faciliter la remise en question des pratiques habituelles. Avant d'envisager une réduction progressive des complications opératoires, il est nécessaire de fiabiliser le processus chirurgical, puis de concevoir des actions d'amélioration spécifiques. Cela implique de mieux comprendre les facteurs qui influencent la performance du chirurgien et le travail en équipe au bloc.

8. Remerciements

Ce travail résulte d'une collaboration étroite entre les équipes du Pôle Information Médicale Évaluation Recherche des Hospices Civils de Lyon et du Service de Chirurgie Générale et Endocrinienne du Centre Hospitalier Lyon Sud.

Nous remercions Jean-Louis Peix, Jean-Christophe Lifante et Pietro Soardo d'avoir accepté de participer activement à cette évaluation de leurs pratiques professionnelles.

Nous remercions Cyrille Colin, Sandrine Touzet et Sandrine Couray-Targe d'avoir apporté leur expertise méthodologique à la conception et la conduite de l'étude.

Nous remercions Nora Gherbi, Joséphine Ros et les internes qui ont permis une collecte rigoureuse des données nécessaires à ce travail, ainsi que l'ensemble des chirurgiens, anesthésistes, biologistes et soignants du Centre Hospitalier Lyon Sud impliqués dans le dépistage des complications opératoires et dont les débats ont animé nos réunions trimestrielles.

Nous remercions l'ensemble des médecins d'information médicale des Hospices Civils de Lyon et les informaticiens du pôle IMER, sans qui nous n'aurions pas disposé de données de qualité, Philippe Messy pour son implication dans leur extraction à partir des bases du PMSI, Nicolas Voirin et Simon Ducarroz pour leur contribution aux analyses statistiques, Thomas Soussy et les documentalistes des Hospices Civils de Lyon pour leur collaboration aux recherches bibliographiques.

Enfin, nous exprimons notre reconnaissance à Bruno Bailly (Haute Autorité de Santé, Saint-Denis), David Biau (Assistance Publique Hôpitaux de Paris), Hélène Duclos (Institut national de la santé et de la recherche médicale, Paris), Jean-Marc Duclos (Fondation Hôpital Saint-Joseph, Paris), Jean-Marie Januel (Université de Lausanne, Suisse), Mohammed Mohammed (University of Birmingham, UK), Robert Lloyd et Lloyd Provost (Insitute for Healthcare Improvement, Boston, USA), Hélène Favre et Marie-Annick Le Pogam (Hospices Civils de Lyon) pour la richesse de nos échanges lors de la conduite de ce travail.

9. Bibliographie

- 1 Bergman S, Feldman LS, Barkun JS. Evaluating surgical outcomes. *Surg Clin North Am*. 2006 Feb;86(1):129-49.
- 2 Feldman L, Barkun J, Barkun A, Sampalis J, Rosenberg L. Measuring postoperative complications in general surgery patients using an outcomes-based strategy: comparison with complications presented at morbidity and mortality rounds. *Surgery*. 1997 Oct;122(4):711-9; discussion 719-20.
- 3 Schimmel EM. The hazards of hospitalization. *Ann Intern Med*. 1964 Jan;60:100-10.
- 4 Steel K, Gertman PM, Crescenzi C, Anderson J. Iatrogenic illness on a general medical service at a university hospital. *N Engl J Med*. 1981 Mar 12;304(11):638-42.
- 5 Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, Newhouse JP, Weiler PC, Hiatt HH. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N Engl J Med*. 1991 Feb 7;324(6):370-6.
- 6 Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, Hebert L, Newhouse JP, Weiler PC, Hiatt H. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. *N Engl J Med*. 1991 Feb 7;324(6):377-84.
- 7 Leape LL, Lawthers AG, Brennan TA, Johnson WG. Preventing medical injury. *QRB Qual Rev Bull*. 1993 May;19(5):144-9.
- 8 Kable AK, Gibberd RW, Spigelman AD. Adverse events in surgical patients in Australia. *Int J Qual Health Care*. 2002 Aug;14(4):269-76.
- 9 Zegers M, de Bruijne MC, Wagner C, Hoonhout LH, Waaijman R, Smits M, Hout FA, Zwaan L, Christiaans-Dingelhoff I, Timmermans DR, Groenewegen PP, van der Wal G. Adverse events and potentially preventable deaths in Dutch hospitals: results of a retrospective patient record review study. *Qual Saf Health Care*. 2009 Aug;18(4):297-302.
- 10 Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. *BMJ*. 2001 Mar 3;322(7285):517-9.
- 11 Soop M, Fryksmark U, Köster M, Haglund B. The incidence of adverse events in Swedish hospitals: a retrospective medical record review study. *Int J Qual Health Care*. 2009 Aug;21(4):285-91.
- 12 de Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA. The incidence and nature of in-hospital adverse events: a systematic review. *Qual Saf Health Care*. 2008 Jun;17(3):216-23.
- 13 Michel P, Quenon JL, Djihoud A, Tricaud-Vialle S, Sarasqueta AM, Domecq S, Haury B, Cases C. Les événements indésirables liés aux soins observés dans les établissements de santé: premiers résultats d'une étude nationale. *DREES, Etudes et Résultats*. Mai 2005;398:1-15.
- 14 Leape LL. Error in medicine. *JAMA*. 1994 Dec 21;272(23):1851-7.
- 15 Pettker CM, Funai EF. Getting it right when things go wrong. *JAMA*. 2010 Mar 10;303(10):977-8.
- 16 Kohn L, ed, Corrigan J, ed, Donaldson M, ed (1999). *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press. (<http://www.iom.edu/Object.File/Master/4/117/ToErr-8pager.pdf>)
- 17 Thomas EJ, Studdert DM, Burstin HR, Orav EJ, Zeena T, Williams EJ, Howard KM, Weiler PC, Brennan TA. Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care*. 2000 Mar;38(3):261-71.
- 18 Hippocrate. *De l'Art Médical*. Epidémies, I, 2, 5 (vers 410 av. J-C).
- 19 Dalton GD, Samaropoulos XF, Dalton AC. Improvements in the safety of patient care can help end the medical malpractice crisis in the United States. *Health Policy*. 2008 May;86(2-3):153-62.

-
- 20 Stelfox HT, Palmisani S, Scurlock C, Orav EJ, Bates DW. The "To Err is Human" report and the patient safety literature. *Qual Saf Health Care*. 2006 Jun;15(3):174-8.
 - 21 Lilford R, Stirling S, Maillard N. Citation classics in patient safety research: an invitation to contribute to an online bibliography. *Qual Saf Health Care*. 2006 Oct;15(5):311-3.
 - 22 Rogers SO Jr. Evaluation of the impact of health services research on quality of care. *Surgery*. 2009 Jun;145(6):635-8.
 - 23 Bates DW. Mountains in the clouds: patient safety research. *Qual Saf Health Care*. 2008 Jun;17(3):156-7.
 - 24 Construction et utilisation des indicateurs dans le domaine de la santé : principes généraux. Agence Nationale d'Accréditation et d'Évaluation en Santé, mai 2002 : 39 p.
 - 25 Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q*. 1966 Jul;44(3):Suppl:166-206.
 - 26 Agency for Healthcare Research and Quality. Child health care quality toolbox: measuring quality in children's health. Understanding quality measurement. Rockville, Maryland, 2004. (<http://www.ahrq.gov/ctoolbx/understn.htm>)
 - 27 Brown C, Hofer T, Johal A, Thomson R, Nicholl J, Franklin BD, Lilford RJ. An epistemology of patient safety research: a framework for study design and interpretation. Part 3. End points and measurement. *Qual Saf Health Care*. 2008 Jun;17(3):170-7.
 - 28 Dudley RA, Johansen KL, Brand R, Rennie DJ, Milstein A. Selective referral to high-volume hospitals: estimating potentially avoidable deaths. *JAMA*. 2000 Mar 1;283(9):1159-66.
 - 29 Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Ann Intern Med*. 2002 Sep 17;137(6):511-20.
 - 30 Gandjour A, Bannenberg A, Lauterbach KW. Threshold volumes associated with higher survival in health care: a systematic review. *Med Care*. 2003 Oct;41(10):1129-41.
 - 31 Killeen SD, O'Sullivan MJ, Coffey JC, Kirwan WO, Redmond HP. Provider volume and outcomes for oncological procedures. *Br J Surg*. 2005 Apr;92(4):389-402.
 - 32 Werner RM, Bradlow ET. Relationship between Medicare's hospital compare performance measures and mortality rates. *JAMA*. 2006 Dec 13;296(22):2694-702.
 - 33 Pitches DW, Mohammed MA, Lilford RJ. What is the empirical evidence that hospitals with higher-risk adjusted mortality rates provide poorer quality care? A systematic review of the literature. *BMC Health Serv Res*. 2007 Jun 20;7:91.
 - 34 Birkmeyer JD, Sun Y, Goldfaden A, Birkmeyer NJ, Stukel TA. Volume and process of care in high-risk cancer surgery. *Cancer*. 2006 Jun 1;106(11):2476-81.
 - 35 Kernisan LP, Lee SJ, Boscardin WJ, Landefeld CS, Dudley RA. Association between hospital-reported Leapfrog Safe Practices Scores and inpatient mortality. *JAMA*. 2009 Apr 1;301(13):1341-8.
 - 36 Iezzoni LI. Using risk-adjusted outcomes to assess clinical practice: an overview of issues pertaining to risk adjustment. *Ann Thorac Surg*. 1994 Dec;58(6):1822-6.
 - 37 Shahian DM, Blackstone EH, Edwards FH, Grover FL, Grunkemeier GL, Naftel DC, Nashef SA, Nugent WC, Peterson ED; STS workforce on evidence-based surgery. Cardiac surgery risk models: a position article. *Ann Thorac Surg*. 2004 Nov;78(5):1868-77.
 - 38 Mant J. Process versus outcome indicators in the assessment of quality of health care. *Int J Qual Health Care*. 2001 Dec;13(6):475-80.
 - 39 Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet*. 2004 Apr 3;363(9415):1147-54.
 - 40 Powell AE, Davies HT, Thomson RG. Using routine comparative data to assess the quality of health care: understanding and avoiding common pitfalls. *Qual Saf Health Care*. 2003 Apr;12(2):122-8.

-
- 41 Miller MR, Elixhauser A, Zhan C, Meyer GS. Patient Safety Indicators: using administrative data to identify potential patient safety concerns. *Health Serv Res.* 2001 Dec;36(6 Pt 2):110-32.
 - 42 Agency for Healthcare Research and Quality. Guide to Patient Safety Indicators. 2006 Feb. (<http://www.qualityindicator.ahrq.gov>)
 - 43 University of California, San Francisco-Stanford University Evidence-Based Practice Center. Evidence Research Measures of Patient Safety Based on Hospital Administrative Data: The Patient Safety Indicators. Rockville, MD: Agency for Healthcare Research and Quality; 2002.
 - 44 Saskia Drösler. Facilitating Cross-National Comparisons of Indicators for Patient Safety at the Health-System Level in the OECD Countries (Health Technical Paper No. 19). OECD Health Technical Papers, 2008. (<http://www.oecd.org/dataoecd/24/48/40401929.pdf>)
 - 45 World Health Organization. The Research Priority Setting Working Group of the World Alliance for Patient Safety. Summary of the Evidence on Patient Safety: Implications for Research. (http://www.who.int/patientsafety/information_centre/20080523_Summary_of_the_evidence_on_patient_safety.pdf)
 - 46 Vincent C, Aylin P, Franklin BD, Holmes A, Iskander S, Jacklin A, Moorthy K. Is health care getting safer? *BMJ.* 2008 Nov 13;337:a2426.
 - 47 Woloshynowych M, Rogers S, Taylor-Adams S, Vincent C. The investigation and analysis of critical incidents and adverse events in healthcare. *Health Technol Assess.* 2005 May;9(19):1-143, iii.
 - 48 Berwick DM, Leape LL. Reducing errors in medicine. *BMJ.* 1999 Jul 17;319(7203):136-7.
 - 49 Department of Health. An organisation with a memory. Report of an Expert Group on Learning from Adverse Events in the NHS. London: The Stationery Office, 2000.
 - 50 Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med.* 2005 Jul 11;165(13):1493-9.
 - 51 Vincent C, Taylor-Adams S, Chapman EJ, Hewett D, Prior S, Strange P, Tizzard A. How to investigate and analyse clinical incidents: clinical risk unit and association of litigation and risk management protocol. *BMJ.* 2000 Mar 18;320(7237):777-81.
 - 52 Reason J. Combating omission errors through task analysis and good reminders. *Qual Saf Health Care.* 2002 Mar;11(1):40-4.
 - 53 Park K. Human error. In : Salvendy G, editor. *Handbook of human factors and ergonomics.* New York : John Wiley and sons ;1997. P. 150-73.
 - 54 Reason JT. *Human error.* New York: Cambridge University Press; 1990.
 - 55 Reason J. Human error: models and management. *BMJ.* 2000 Mar 18;320(7237):768-70.
 - 56 Brown C, Hofer T, Johal A, Thomson R, Nicholl J, Franklin BD, Lilford RJ. An epistemology of patient safety research: a framework for study design and interpretation. Part 1. Conceptualising and developing interventions. *Qual Saf Health Care.* 2008 Jun;17(3):158-62.
 - 57 Amalberti R, Gremion C, Auroy Y, Michel P, Salmi R, Parneix P, Pouchadon ML, Hoarau H, Occelli P, Quenon JL, Hubert B. Les systèmes de signalement des événements indésirables en médecine. *DREES, Etudes et Résultats.* Juillet 2007;584:1-7.
 - 58 Vincent C, Taylor-Adams S, Stanhope N. Framework for analysing risk and safety in clinical medicine. *BMJ.* 1998 Apr 11;316(7138):1154-7.
 - 59 Weiser TG, Regenbogen SE, Thompson KD, Haynes AB, Lipsitz SR, Berry WR, Gawande AA. An estimation of the global volume of surgery: a modelling strategy based on available data. *Lancet.* 2008 Jul 12;372(9633):139-44.
 - 60 Statistiques hospitalières en soins de courte durée, extraites du Programme de Médicalisation des Systèmes d'Information sur les bases nationales publique et privée de l'année 2007. (<http://www.atih.sante.fr/>)
 - 61 World Health Organisation ; Safe Surgery Saves Lives : The Second Global Patient Safety Challenge. (<http://www.who.int/patientsafety/safesurgery/en/>)

- 62 Gawande AA, Thomas EJ, Zinner MJ, Brennan TA. The incidence and nature of surgical adverse events in Colorado and Utah in 1992. *Surgery*. 1999 Jul;126(1):66-75.
- 63 Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in hospital mortality associated with inpatient surgery. *N Engl J Med*. 2009 Oct 1;361(14):1368-75.
- 64 Weiser TG, Makary MA, Haynes AB, Dziekan G, Berry WR, Gawande AA; Safe Surgery Saves Lives Measurement and Study Groups. Standardised metrics for global surgical surveillance. *Lancet*. 2009 Sep 26;374(9695):1113-7.
- 65 Haynes AB, Weiser TG, Berry WR, Lipsitz SR, Breizat AH, Dellinger EP, Herbosa T, Joseph S, Kibatala PL, Lapitan MC, Merry AF, Moorthy K, Reznick RK, Taylor B, Gawande AA; Safe Surgery Saves Lives Study Group. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med*. 2009 Jan 29;360(5):491-9.
- 66 Haute Autorité de Santé, La check-list « Sécurité du patient au bloc opératoire », http://www.has-sante.fr/portail/jcms/c_821871/la-check-list-securite-du-patient-au-bloc-operatoire
- 67 Ergina PL, Cook JA, Blazeby JM, Boutron I, Clavien PA, Reeves BC, Seiler CM; Balliol Collaboration, Altman DG, Aronson JK, Barkun JS, Campbell WB, Cook JA, Feldman LS, Flum DR, Glasziou P, Maddern GJ, Marshall JC, McCulloch P, Nicholl J, Strasberg SM, Meakins JL, Ashby D, Black N, Bunker J, Burton M, Campbell M, Chalkidou K, Chalmers I, de Leval M, Deeks J, Grant A, Gray M, Greenhalgh R, Jenicek M, Kehoe S, Lilford R, Littlejohns P, Loke Y, Madhock R, McPherson K, Rothwell P, Summerskill B, Taggart D, Tekkis P, Thompson M, Treasure T, Trohler U, Vandenbroucke J. Challenges in evaluating surgical innovation. *Lancet*. 2009 Sep 26;374(9695):1097-104.
- 68 Bratzler DW, Houck PM, Richards C, Steele L, Dellinger EP, Fry DE, Wright C, Ma A, Carr K, Red L. Use of antimicrobial prophylaxis for major surgery: baseline results from the National Surgical Infection Prevention Project. *Arch Surg*. 2005 Feb;140(2):174-82.
- 69 Howie C, Hughes H, Watts AC. Venous thromboembolism associated with hip and knee replacement over a ten-year period: a population-based study. *J Bone Joint Surg Br*. 2005 Dec;87(12):1675-80.
- 70 Kwaan MR, Studdert DM, Zinner MJ, Gawande AA. Incidence, patterns, and prevention of wrong-site surgery. *Arch Surg*. 2006 Apr;141(4):353-7.
- 71 Gawande AA, Studdert DM, Orav EJ, Brennan TA, Zinner MJ. Risk factors for retained instruments and sponges after surgery. *N Engl J Med*. 2003 Jan 16;348(3):229-35.
- 72 Bruce J, Russell EM, Mollison J, Krukowski ZH. The measurement and monitoring of surgical adverse events. *Health Technol Assess*. 2001;5(22):1-194.
- 73 Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. *Br J Surg*. 1991 Mar;78(3):355-60.
- 74 Carter D. The surgeon as a risk factor. *BMJ*. 2003 Apr 19;326(7394):832-3.
- 75 Sosa JA, Bowman HM, Tielsch JM, Powe NR, Gordon TA, Udelsman R. The importance of surgeon experience for clinical and economic outcomes from thyroidectomy. *Ann Surg*. 1998 Sep;228(3):320-30.
- 76 Luft HS, Bunker JP, Enthoven AC. Should operations be regionalized? The empirical relation between surgical volume and mortality. *N Engl J Med*. 1979 Dec 20;301(25):1364-9.
- 77 Hannan EL, O'Donnell JF, Kilburn H Jr, Bernard HR, Yazici A. Investigation of the relationship between volume and mortality for surgical procedures performed in New York State hospitals. *JAMA*. 1989 Jul 28;262(4):503-10.
- 78 Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT. Statistical assessment of the learning curves of health technologies. *Health Technol Assess*. 2001;5(12):1-79.
- 79 Vickers AJ, Bianco FJ, Serio AM, Eastham JA, Schrag D, Klein EA, Reuther AM, Kattan MW, Pontes JE, Scardino PT. The surgical learning curve for prostate cancer control after radical prostatectomy. *J Natl Cancer Inst*. 2007 Aug 1;99(15):1171-7.

-
- 80 Vincent C, Moorthy K, Sarker SK, Chang A, Darzi AW. Systems approaches to surgical quality and safety: from concept to measurement. *Ann Surg.* 2004 Apr;239(4):475-82.
- 81 Smith L. Working time, stress and fatigue. In: Vincent CA, ed. *Clinical Risk Management. Enhancing Patient Safety*, London: BMJ Publications; 2001.
- 82 Keyes C. "Every defect is a treasure". *Int J Qual Health Care.* 1997 Oct;9(5):391-2.
- 83 Rosenstein AH, O'Daniel M. Impact and implications of disruptive behavior in the perioperative arena. *J Am Coll Surg.* 2006 Jul;203(1):96-105.
- 84 Undre S, Healey AN, Darzi A, Vincent CA. Observational assessment of surgical teamwork: a feasibility study. *World J Surg.* 2006 Oct;30(10):1774-83.
- 85 Burke CS, Salas E, Wilson-Donnelly K, Priest H. How to turn a team of experts into an expert medical team: guidance from the aviation and military communities. *Qual Saf Health Care.* 2004 Oct;13 Suppl 1:i96-104.
- 86 Makary MA, Sexton JB, Freischlag JA, Millman EA, Pryor D, Holzmüller C, Pronovost PJ. Patient safety in surgery. *Ann Surg.* 2006 May;243(5):628-32; discussion 632-5.
- 87 Gawande AA, Zinner MJ, Studdert DM, Brennan TA. Analysis of errors reported by surgeons at three teaching hospitals. *Surgery.* 2003 Jun;133(6):614-21.
- 88 Dankelman J, Grimbergen CA. Systems approach to reduce errors in surgery. *Surg Endosc.* 2005 Aug;19(8):1017-21.
- 89 Walter A Shewhart dans Wikipedia (http://fr.wikipedia.org/wiki/Walter_A._Shewhart)
- 90 Bailly B, Durand-Zaleski I, De Vernejoul N. Maîtrise statistique des processus en santé : comprendre et expérimenter. Haute Autorité de Santé, Service évaluation des pratiques, décembre 2004 : 92 p.
- 91 Shewhart WA. *Economic control of quality of manufactured product*. New York: Van Nostrand, 1931.
- 92 Berwick DM. Controlling variation in health care: a consultation from Walter Shewhart. *Med Care.* 1991 Dec;29(12):1212-25.
- 93 Shewhart WA. *Statistical Method from the Viewpoint of Quality Control*. Washington DC: Graduate School of the Department of Agriculture, 1939.
- 94 Best M, Neuhauser D. Walter A Shewhart, 1924, and the Hawthorne factory. *Qual Saf Health Care.* 2006 Apr;15(2):142-3.
- 95 Shewhart WA, Deming WE. *Statistical method from the viewpoint of quality control*. 2nd ed. New York: Dover, 1986.
- 96 Montgomery, Douglas C. *Statistical Quality Control: A Modern Introduction*, International Student Version, 6th ed. Hoboken, NJ: Wiley, 2008.
- 97 Mohammed MA, Worthington P, Woodall WH. Plotting basic control charts:tutorial notes for healthcare practitioners. *Qual Saf Health Care.* 2008 Apr;17(2):137-45.
- 98 Marshall T, Mohammed MA. Differences in clinical performance. *Br J Surg.* 2002 Aug;89(8):948-9.
- 99 Benneyan JC, Lloyd RC, Plsek PE. Statistical process control as a tool for research and healthcare improvement. *Qual Saf Health Care.* 2003 Dec;12(6):458-64.
- 100 Ryan, Thomas P. *Statistical methods for quality improvement*, 2nd ed. New York: Wiley, 2000.
- 101 Deming WE. On a classification of the problems of statistical inference. *J Am Stat Ass* 1942; 37(218): 173-85.
- 102 Berwick D, Godfrey AB, Roessner J. *Curing health care: new strategies for quality improvement: a report on the National Demonstration Project on Quality Improvement in Health Care*. San Francisco: Jossey-Bass; 1990.

-
- 103 Joint Commission on Accreditation of Healthcare Organizations. Looking at the ORYX Data 2004.
(<http://www.jcaho.org/accredited+organizations/hospitals/oryx/oryx+survey+process/index.htm>)
- 104 Benneyan JC. Performance of number-between g-type statistical control charts for monitoring adverse events. *Health Care Manag Sci*. 2001 Dec;4(4):319-36.
- 105 Thor J, Lundberg J, Ask J, Olsson J, Carli C, Härenstam KP, Brommels M. Application of statistical process control in healthcare improvement: systematic review. *Qual Saf Health Care*. 2007 Oct;16(5):387-99.
- 106 Mohammed MA, Cheng KK, Rouse A, Marshall T. Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons. *Lancet*. 2001 Feb 10;357(9254):463-7.
- 107 Norberg A, Christopher NC, Ramundo ML, Bower JR, Berman SA. Contamination rates of blood cultures obtained by dedicated phlebotomy vs intravenous catheter. *JAMA*. 2003 Feb 12;289(6):726-9.
- 108 Diaz M, Neuhauser D. Pasteur and parachutes: when statistical process control is better than a randomized controlled trial. *Qual Saf Health Care*. 2005 Apr;14(2):140-3.
- 109 Morton AP, Whitby M, McLaws ML, Dobson A, McElwain S, Looke D, Stackelroth J, Sartor A. The application of statistical process control charts to the detection and monitoring of hospital-acquired infections. *J Qual Clin Pract*. 2001 Dec;21(4):112-7.
- 110 Hanslik T, Boelle PY, Flahault A. The control chart: an epidemiological tool for public health monitoring. *Public Health*. 2001 Jul;115(4):277-81.
- 111 Tennant R, Mohammed MA, Coleman JJ, Martin U. Monitoring patients using control charts: a systematic review. *Int J Qual Health Care*. 2007 Aug;19(4):187-94.
- 112 Marshall T, Mohammed MA. Case-mix and the use of control charts in monitoring mortality rates after coronary artery bypass. *BMC Health Serv Res*. 2007 Apr 30;7:63.
- 113 Hart MK, Lee KY, Hart RF, Robertson JW. Application of attribute control charts to risk-adjusted data for monitoring and improving health care performance. *Qual Manag Health Care*. 2003 Jan-Mar;12(1):5-19.
- 114 Alemi F, Oliver DW. Tutorial on risk-adjusted P-charts. *Qual Manag Health Care*. 2001 Fall;10(1):1-9.
- 115 Woodall WH. The use of control charts in health-care and public-health surveillance, *J Qual Technol* 2006; 38(2): 89-104.
- 116 Sibanda T, Sibanda N. The CUSUM chart method as a tool for continuous monitoring of clinical outcomes using routinely collected data. *BMC Med Res Methodol*. 2007 Nov 3;7:46.
- 117 Biau DJ, Resche-Rigon M, Godiris-Petit G, Nizard RS, Porcher R. Quality control of surgical and interventional procedures: a review of the CUSUM. *Qual Saf Health Care*. 2007 Jun;16(3):203-7.
- 118 Yap CH, Colson ME, Watters DA. Cumulative sum techniques for surgeons: a brief review. *ANZ J Surg*. 2007 Jul;77(7):583-6.
- 119 Lim TO, Soraya A, Ding LM, Morad Z. Assessing doctors' competence: application of CUSUM technique in monitoring doctors' performance. *Int J Qual Health Care*. 2002 Jun;14(3):251-8.
- 120 Bolsin S, Colson M. The use of the Cusum technique in the assessment of trainee competence in new procedures. *Int J Qual Health Care*. 2000 Oct;12(5):433-8.
- 121 Hawkins DM, Olwell DH. Cumulative Sum Charts and Charting for Quality Improvement. Springer Verlag, New York., 1998.
- 122 Rogers CA, Reeves BC, Caputo M, Ganesh JS, Bonser RS, Angelini GD. Control chart methods for monitoring cardiac surgical performance and their interpretation. *J Thorac Cardiovasc Surg*. 2004 Dec;128(6):811-9.
- 123 Lovegrove J, Valencia O, Treasure T, Sherlaw-Johnson C, Gallivan S. Monitoring the results of cardiac surgery by variable life-adjusted display. *Lancet*. 1997 Oct 18;350(9085):1128-30.

-
- 124 Poloniecki J, Valencia O, Littlejohns P. Cumulative risk adjusted mortality chart for detecting changes in death rate: observational study of heart surgery. *BMJ*. 1998 Jun 6;316(7146):1697-700. Erratum in: *BMJ* 1998 Jun 27;316(7149):1947.
 - 125 Wald A. *Sequential Analysis*. John Wiley & Sons, New York., 1947.
 - 126 Grunkemeier GL, Jin R, Wu Y. Cumulative sum curves and their prediction limits. *Ann Thorac Surg*. 2009 Feb;87(2):361-4.
 - 127 Spiegelhalter D, Grigg O, Kinsman R, Treasure T. Risk-adjusted sequential probability ratio tests: applications to Bristol, Shipman and adult cardiac surgery. *Int J Qual Health Care*. 2003 Feb;15(1):7-13.
 - 128 Page ES. Continuous inspection schemes. *Biometrika*. 1954;41:10014.
 - 129 de Leval MR, François K, Bull C, Brawn W, Spiegelhalter D. Analysis of a cluster of surgical failures. Application to a series of neonatal arterial switch operations. *J Thorac Cardiovasc Surg*. 1994 Mar;107(3):914-23; discussion 923-4.
 - 130 Steiner SH, Cook RJ, Farewell VT, Treasure T. Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics*. 2000 Dec;1(4):441-52.
 - 131 Biau DJ, Williams SM, Schlup MM, Nizard RS, Porcher R. Quantitative and individualized assessment of the learning curve using LC-CUSUM. *Br J Surg*. 2008 Jul;95(7):925-9.
 - 132 Neubauer AS. The EWMA control chart: properties and comparison with other quality-control procedures by computer simulation. *Clin Chem*. 1997 Apr;43(4):594-601.
 - 133 Woodall WH. Controversies and contradictions in statistical process control. *J Qual Technol*. 2000 Oct;32(4):41-350.
 - 134 Adab P, Rouse AM, Mohammed MA, Marshall T. Performance league tables: the NHS deserves better. *BMJ*. 2002 Jan 12;324(7329):95-8.
 - 135 Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Stat Med*. 2005 Apr 30;24(8):1185-202.
 - 136 Spiegelhalter DJ. Handling over-dispersion of performance indicators. *Qual Saf Health Care*. 2005 Oct;14(5):347-51.
 - 137 Bliss RD, Gauger PG, Delbridge LW. Surgeon's approach to the thyroid gland: surgical anatomy and the importance of technique. *World J Surg*. 2000 Aug;24(8):891-7.
 - 138 Akerström G, Malmaeus J, Bergström R. Surgical anatomy of human parathyroid glands. *Surgery*. 1984 Jan;95(1):14-21.
 - 139 Chanson P. Hypothyroidism in adults. *Rev Prat*. 2006 Jun 30;56(12):1369-75.
 - 140 Association Francophone de Chirurgie Endocrinienne. Note d'information au patient. (http://www.chirurgie-endocrinienne.net/download/afce_livret_info_thyroide.pdf)
 - 141 Fondation Canadienne de la Thyroïde. Thyroid diseases: what you need to know. *Rev Prat*. 2005 Jan 31;55(2):194.
 - 142 Surks MI, Ortiz E, Daniels GH, Sawin CT, Col NF, Cobin RH, Franklyn JA, Hershman JM, Burman KD, Denke MA, Gorman C, Cooper RS, Weissman NJ. Subclinical thyroid disease: scientific review and guidelines for diagnosis and management. *JAMA*. 2004 Jan 14;291(2):228-38.
 - 143 Lind P, Langsteger W, Molnar M, Gallowitsch HJ, Mikosch P, Gomez I. Epidemiology of thyroid diseases in iodine sufficiency. *Thyroid*. 1998 Dec;8(12):1179-83.
 - 144 Zulewski H, Müller B, Exer P, Miserez AR, Staub JJ. Estimation of tissue hypothyroidism by a new clinical score: evaluation of patients with various grades of hypothyroidism and controls *J Clin Endocrinol Metab*. 1997 Mar;82(3):771-6.
 - 145 Wemeau JL. Hyperthyroidism: current concepts and management. *Rev Prat*. 2005 Jan 31;55(2):149-57.
 - 146 Sherman SI. Thyroid carcinoma. *Lancet*. 2003 Feb 8;361(9356):501-11.

- 147 Remontet L, Estève J, Bouvier AM, Grosclaude P, Launoy G, Menegoz F, Exbrayat C, Tretare B, Carli PM, Guizard AV, Troussard X, Bercelli P, Colonna M, Halna JM, Hedelin G, Macé-Lesec'h J, Peng J, Buemi A, Velten M, Jouglu E, Arveux P, Le Bodic L, Michel E, Sauvage M, Schvartz C, Faivre J. Cancer incidence and mortality in France over the period 1978-2000. *Rev Epidemiol Sante Publique*. 2003 Feb;51(1 Pt 1):3-30.
- 148 Davies L, Welch HG. Increasing incidence of thyroid cancer in the United States, 1973-2002. *JAMA*. 2006 May 10;295(18):2164-7.
- 149 Hayat MJ, Howlader N, Reichman ME, Edwards BK. Cancer statistics, trends, and multiple primary cancer analyses from the Surveillance, Epidemiology, and End Results (SEER) Program. *Oncologist*. 2007 Jan;12(1):20-37.
- 150 Todesco J, Williams RT. Anaesthetic management of a patient with a large neck mass. *Can J Anaesth*. 1994 Feb;41(2):157-60.
- 151 Peix JL. "Thyroidectomy" becoming excessive surgery? *Ann Chir*. 2002 Feb;127(2):85-7.
- 152 Klein M, Brunaud L. Goiter and thyroid nodules. *Rev Prat*. 2008 Jun 15;58(11):1249-59.
- 153 Gough IR, Wilkinson D. Total thyroidectomy for management of thyroid disease. *World J Surg*. 2000 Aug;24(8):962-5.
- 154 National Comprehensive Cancer Network (NCCN). Clinical Practice Guidelines in Oncology: Thyroid Carcinoma V.I.2008. 2008. (http://www.nccn.org/professionals/physician_gls/PDF/thyroid.pdf)
- 155 Welbourn RB. The History of Endocrine Surgery, New York, Praeger, 1990 : pp. 19–82.
- 156 Greene WW. Three cases of bronchocele successfully removed. *Am J Med Sci*. 1871;121:80.
- 157 Vellar ID. Thomas Peel Dunhill: pioneer thyroid surgeon. *Aust N Z J Surg*. 1999 May;69(5):375-87.
- 158 Gimm O, Brauckhoff M, Thanh PN, Sekulla C, Dralle H. An update on thyroid surgery. *Eur J Nucl Med Mol Imaging*. 2002 Aug;29 Suppl 2:S447-52.
- 159 Thompson NW, Olsen WR, Hoffman GL. The continuing development of the technique of thyroidectomy. *Surgery*. 1973 Jun;73(6):913-27.
- 160 Bérard L, Peycelon R. Traitement chirurgical de la maladie de Basedow et des goîtres toxiques. Paris: Masson, 1936.
- 161 Lahey FH, Hoover WB. Injuries to the recurrent laryngeal nerve in thyroid operations : their management and avoidance. *Ann Surg*. 1938 Oct;108(4):545-62.
- 162 Jatzko GR, Lisborg PH, Müller MG, Wette VM. Recurrent nerve palsy after thyroid operations--principal nerve identification and a literature review. *Surgery*. 1994 Feb;115(2):139-44.
- 163 Hermann M, Alk G, Roka R, Glaser K, Freissmuth M. Laryngeal recurrent nerve injury in surgery for benign thyroid diseases: effect of nerve dissection and impact of individual surgeon in more than 27,000 nerves at risk. *Ann Surg*. 2002 Feb;235(2):261-8.
- 164 Terris DJ, Stack BC Jr. Current technology in thyroid surgery. *ORL J Otorhinolaryngol Relat Spec*. 2008;70(5):305-12.
- 165 Shedd DP, Burget GC. Identification of the recurrent laryngeal nerve. *Arch Surg*. 1966 Jun;92(6):861-4.
- 166 Sturgeon C, Sturgeon T, Angelos P. Neuromonitoring in thyroid surgery: attitudes, usage patterns, and predictors of use among endocrine surgeons. *World J Surg*. 2009 Mar;33(3):417-25.
- 167 Dralle H, Sekulla C, Lorenz K, Brauckhoff M, Machens A; German IONM Study Group. Intraoperative monitoring of the recurrent laryngeal nerve in thyroid surgery. *World J Surg*. 2008 Jul;32(7):1358-66.
- 168 Barczyński M, Konturek A, Cichoń S. Randomized clinical trial of visualization versus neuromonitoring of recurrent laryngeal nerves during thyroidectomy. *Br J Surg*. 2009 Mar;96(3):240-6.

-
- 169 Saxe AW, Brown E, Hamburger SW. Thyroid and parathyroid surgery performed with patient under regional anesthesia. *Surgery*. 1988 Apr;103(4):415-20.
- 170 Lo Gerfo P, Gates R, Gazetas P. Outpatient and short-stay thyroid surgery. *Head Neck*. 1991 Mar-Apr;13(2):97-101.
- 171 Voutilainen PE, Haapiainen RK, Haglund CH. Ultrasonically activated shears in thyroid surgery. *Am J Surg*. 1998 Jun;175(6):491-3.
- 172 Miccoli P, Berti P, Materazzi G, Minuto M, Barellini L. Minimally invasive video-assisted thyroidectomy: five years of experience. *J Am Coll Surg*. 2004 Aug;199(2):243-8.
- 173 Littmann K, Magdsick G, Strötges MW, Eigler FW. Intraoperative localization measurement following preoperative radioiodine marking to facilitate the treatment of differentiated thyroid carcinoma. *Chirurg*. 1980 Jun;51(6):389-94.
- 174 Abboud B, Sleilaty G, Zeineddine S, Braidy C, Aouad R, Tohme C, Noun R, Sarkis R. Is therapy with calcium and vitamin D and parathyroid autotransplantation useful in total thyroidectomy for preventing hypocalcemia? *Head Neck*. 2008 Sep;30(9):1148-54.
- 175 Hegner CF. A History of Thyroid Surgery. *Ann Surg*. 1932 Apr;95(4):481-92.
- 176 Trésallet C, Chigot JP, Menegaux F. How to prevent recurrent nerve palsy during thyroid surgery?. *Ann Chir*. 2006 Feb;131(2):149-53.
- 177 Farling PA. Thyroid disease. *Br J Anaesth*. 2000 Jul;85(1):15-28.
- 178 Burge MR, Zeise TM, Johnsen MW, Conway MJ, Qualls CR. Risks of complication following thyroidectomy. *J Gen Intern Med*. 1998 Jan;13(1):24-31.
- 179 Chiang FY, Lu IC, Kuo WR, Lee KW, Chang NC, Wu CW. The mechanism of recurrent laryngeal nerve injury during thyroid surgery--the application of intraoperative neuromonitoring. *Surgery*. 2008 Jun;143(6):743-9.
- 180 Spector BC, Netterville JL, Billante C, Clary J, Reinisch L, Smith TL. Quality-of-life assessment in patients with unilateral vocal cord paralysis. *Otolaryngol Head Neck Surg*. 2001 Sep;125(3):176-82.
- 181 Munch S, deKryger L. A piece of my mind. Moral wounds: complicated complications. *JAMA*. 2001 Mar 7;285(9):1131-2.
- 182 Kern KA. Medicolegal analysis of errors in diagnosis and treatment of surgical endocrine disease. *Surgery*. 1993 Dec;114(6):1167-73; discussion 1173-4.
- 183 Shaw GY, Pierce E. Malpractice litigation involving iatrogenic surgical vocal fold paralysis: a closed-claims review with recommendations for prevention and management. *Ann Otol Rhinol Laryngol* 2009; 118(1): 6-12.
- 184 Thomusch O, Machens A, Sekulla C, Ukkat J, Lippert H, Gastinger I, Dralle H. Multivariate analysis of risk factors for postoperative complications in benign goiter surgery: prospective multicenter study in Germany. *World J Surg*. 2000 Nov;24(11):1335-41.
- 185 Dralle H, Sekulla C, Haerting J, Timmermann W, Neumann HJ, Kruse E, Grond S, Mühlig HP, Richter C, Voss J, Thomusch O, Lippert H, Gastinger I, Brauckhoff M, Gimm O. Risk factors of paralysis and functional outcome after recurrent laryngeal nerve monitoring in thyroid surgery. *Surgery*. 2004 Dec;136(6):1310-22.
- 186 Thomusch O, Machens A, Sekulla C, Ukkat J, Brauckhoff M, Dralle H. The impact of surgical technique on postoperative hypoparathyroidism in bilateral thyroid surgery: a multivariate analysis of 5846 consecutive patients. *Surgery*. 2003 Feb;133(2):180-5.
- 187 Dimich A, Bedrossian P, Wallach S. Hypoparathyroidism: clinical observations in 34 patients. *Arch Intern Med*. 1967; 120:449-58.
- 188 Affleck BD, Swartz K, Brennan J. Surgical considerations and controversies in thyroid and parathyroid surgery. *Otolaryngol Clin North Am*. 2003 Feb;36(1):159-87, x.
- 189 Hundahl SA, Cady B, Cunningham MP, Mazzaferri E, McKee RF, Rosai J, Shah JP, Fremgen AM, Stewart AK, Hölzer S. Initial results from a prospective cohort study of 5583 cases of

-
- thyroid carcinoma treated in the united states during 1996. U.S. and German Thyroid Cancer Study Group. An American College of Surgeons Commission on Cancer Patient Care Evaluation study. *Cancer*. 2000 Jul 1;89(1):202-17.
- 190 Clinical Practice Guidelines. Recurrent Laryngeal Nerve Paralysis in Adults. Octobre 2002. *Annales d'Otolaryngologie et de Chirurgie Cervico-Faciale* juin 2003; 120 (3):182-199. (http://www.orl-france.org/download/RPC3_Texte_court.pdf)
- 191 (http://www.chirurgie-endocrinienne.net/download/afce_livret_info_thyroide.pdf)
- 192 (<http://www.baes.info/Pages/BAETS%20Guidelines.pdf>)
- 193 (<http://www.rcplondon.ac.uk/pubs/contents/5e0be7d7-0726-48af-a510-5cc5b9ec574d.pdf>)
- 194 AES GUIDELINES 06/01 GROUP. Australian endocrine surgeons guidelines AES06/01. Postoperative parathyroid hormone measurement and early discharge after total thyroidectomy: analysis of Australian data and management recommendations. *ANZ J. Surg.* 2007; 77: 199–202.
- 195 Montori VM, Wilczynski NL, Morgan D, Haynes RB; Hedges Team. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *BMJ*. 2005 Jan 8;330(7482):68.
- 196 Safioleas M, Stamatakis M, Rompoti N, Mouzopoulos G, Iannescu R, Salichou V, Skandalakis P. Complications of thyroid surgery. *Chirurgia (Bucur)*. 2006 Nov-Dec;101(6):571-81.
- 197 Fewins J, Simpson CB, Miller FR. Complications of thyroid and parathyroid surgery. *Otolaryngol Clin North Am*. 2003 Feb;36(1):189-206, x.
- 198 Netterville JL, Aly A, Ossoff RH. Evaluation and treatment of complications of thyroid and parathyroid surgery. *Otolaryngol Clin North Am*. 1990 Jun;23(3):529-52.
- 199 Dionigi G, Bacuzzi A, Bertocchi V, Carrafiello G, Boni L, Rovera F, Dionigi R. Safe incorporation of new technologies in thyroid surgery. *Expert Rev Med Devices*. 2008 Nov;5(6):747-58.
- 200 Noordzij JP, Lee SL, Bernet VJ, Payne RJ, Cohen SM, McLeod IK, Hier MP, Black MJ, Kerr PD, Richards ML, Lo CY, Raffaelli M, Bellantone R, Lombardi CP, Cohen JI, Dietrich MS. Early prediction of hypocalcemia after thyroidectomy using parathyroid hormone: an analysis of pooled individual patient data from nine observational studies. *J Am Coll Surg*. 2007 Dec;205(6):748-54.
- 201 Grodski S, Serpell J. Evidence for the role of perioperative PTH measurement after total thyroidectomy as a predictor of hypocalcemia. *World J Surg*. 2008 Jul;32(7):1367-73.
- 202 Rosato L, Avenia N, Bernante P, De Palma M, Gulino G, Nasi PG, Pelizzo MR, Pezzullo L. Complications of thyroid surgery: analysis of a multicentric study on 14,934 patients operated on in Italy over 5 years. *World J Surg*. 2004 Mar;28(3):271-6.
- 203 Pattou F, Combemale F, Fabre S, Carnaille B, Decoux M, Wemeau JL, Racadot A, Proye C. Hypocalcemia following thyroid surgery: incidence and prediction of outcome. *World J Surg*. 1998 Jul;22(7):718-24.
- 204 Jeannon JP, Orabi AA, Bruch GA, Abdalsalam HA, Simo R. Diagnosis of recurrent laryngeal nerve palsy after thyroidectomy: a systematic review. *Int J Clin Pract*. 2009 Apr;63(4):624-9.
- 205 Mishra A, Agarwal G, Agarwal A, Mishra SK. Safety and efficacy of total thyroidectomy in hands of endocrine surgery trainees. *Am J Surg*. 1999 Nov;178(5):377-80.
- 206 Manolidis S, Takashima M, Kirby M, Scarlett M. Thyroid surgery: a comparison of outcomes between experts and surgeons in training. *Otolaryngol Head Neck Surg*. 2001 Jul;125(1):30-3.
- 207 Acun Z, Cihan A, Ulukent SC, Comert M, Ucan B, Cakmak GK, Cesur A. A randomized prospective study of complications between general surgery residents and attending surgeons in near-total thyroidectomies. *Surg Today*. 2004;34(12):997-1001.

-
- 208 Hassan I, Koller M, Kluge C, Hoffmann S, Zielke A, Rothmund M. Supervised surgical trainees perform thyroid surgery for Graves' disease safely. *Langenbecks Arch Surg.* 2006 Nov;391(6):597-602.
- 209 Reeve TS, Curtin A, Fingleton L, Kennedy P, Mackie W, Porter T, Simons D, Townend D, Delbridge L. Can total thyroidectomy be performed as safely by general surgeons in provincial centers as by surgeons in specialized endocrine surgical units? Making the case for surgical training. *Arch Surg.* 1994 Aug;129(8):834-6.
- 210 Sosa JA, Bowman HM, Tielsch JM, Powe NR, Gordon TA, Udelsman R. The importance of surgeon experience for clinical and economic outcomes from thyroidectomy. *Ann Surg.* 1998 Sep;228(3):320-30.
- 211 Stavrakis AI, Ituarte PH, Ko CY, Yeh MW. Surgeon volume as a predictor of outcomes in inpatient and outpatient endocrine surgery. *Surgery.* 2007 Dec;142(6):887-99.
- 212 Pieracci FM, Fahey TJ 3rd. Effect of hospital volume of thyroidectomies on outcomes following substernal thyroidectomy. *World J Surg.* 2008 May;32(5):740-6.
- 213 Lifante JC, Duclos A, Couray-Targe S, Colin C, Peix JL, Schott AM. Hospital volume influences the choice of operation for thyroid cancer. *Br J Surg.* 2009 Nov;96(11):1284-8.
- 214 Aylin P, Bottle A, Majeed A. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. *BMJ.* 2007 May 19;334(7602):1044.
- 215 Hosmer DW, Lemeshow S. *Applied Logistic Regression.* New York: Wiley 1989.
- 216 Fasting S, Gisvold SE. Statistical process control methods allow the analysis and improvement of anesthesia care. *Can J Anaesth.* 2003 Oct;50(8):767-74.
- 217 Pollard JB, Garnerin P. Outpatient preoperative evaluation clinic can lead to a rapid shift from inpatient to outpatient surgery: a retrospective review of perioperative setting and outcome. *J Clin Anesth.* 1999 Feb;11(1):39-45.
- 218 Böelle PY, Bonnet F, Valleron AJ. An integrated system for significant anaesthetic events monitoring. *Proc AMIA Symp.* 2001:62-5.
- 219 Levett JM, Carey RG. Measuring for improvement: from Toyota to thoracic surgery. *Ann Thorac Surg.* 1999 Aug;68(2):353-8; discussion 374-6.
- 220 Shahian DM, Williamson WA, Svensson LG, Restuccia JD, D'Agostino RS. Applications of statistical quality control to cardiac surgery. *Ann Thorac Surg.* 1996 Nov;62(5):1351-8; discussion 1358-9.
- 221 Biau DJ, Milet A, Thévenin F, Anract P, Porcher R. Monitoring surgical performance: an application to total hip replacement. *J Eval Clin Pract.* 2009 Jun;15(3):420-4.
- 222 Konrad C, Gerber H, Schupfer G, Jenzer S, Schmucki O. Detection of fluid volume absorption by end-tidal alcohol monitoring in patients undergoing endoscopic renal pelvic surgery. *J Clin Anesth.* 1999 Aug;11(5):386-90.
- 223 Macé L, Bertrand S, Lucron H, Grollmuss O, Dopff C, Mattéi MF, Bossier G, Le Tacon S, Monin P, Marçon F. Paediatric cardiac surgery and autoevaluation: risk score, complexity score and graphic analysis. *Arch Mal Coeur Vaiss.* 2005 May;98(5):477-84.
- 224 Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev.* 2006 Apr 19;(2):CD000259.
- 225 Werner RM, McNutt R. A new strategy to improve quality: rewarding actions rather than measures. *JAMA.* 2009 Apr 1;301(13):1375-7.
- 226 Reznick RK, MacRae H. Teaching surgical skills--changes in the wind. *N Engl J Med.* 2006 Dec 21;355(25):2664-9.
- 227 Paradis C. Bias in surgical research. *Ann Surg.* 2008 Aug;248(2):180-8.
- 228 Cook JA. The challenges faced in the design, conduct and analysis of surgical randomised controlled trials. *Trials.* 2009 Feb 6;10:9.

-
- 229 Eccles M, Grimshaw J, Campbell M, Ramsay C. Research designs for studies evaluating the effectiveness of change and improvement strategies. *Qual Saf Health Care*. 2003 Feb;12(1):47-52.
- 230 Ravaud P, Durieux P. Guide méthodologique : Méthodes quantitatives pour évaluer les interventions visant à améliorer les pratiques. Haute Autorité de santé, Service des recommandations professionnelles, juin 2007 : 59 p.
- 231 Couray-Targe S, Ecochard R, Colin C. Program for medicalization of the information systems in France: tool for management and quality assessment? *Presse Med*. 1999 Oct 2;28(29):1597-603.
- 232 Staiger DO, Dimick JB, Baser O, Fan Z, Birkmeyer JD. Empirically derived composite measures of surgical performance. *Med Care*. 2009 Feb;47(2):226-33.
- 233 Khuri SF, Daley J, Henderson W, Barbour G, Lowry P, Irvin G, Gibbs J, Grover F, Hammermeister K, Strempel JF, et al. The National Veterans Administration Surgical Risk Study: risk adjustment for the comparative assessment of the quality of surgical care. *J Am Coll Surg*. 1995 May;180(5):519-31.
- 234 Romano PS, Mull HJ, Rivard PE, Zhao S, Henderson WG, Loveland S, Tsilimingras D, Christiansen CL, Rosen AK. Validity of selected AHRQ patient safety indicators based on VA National Surgical Quality Improvement Program data. *Health Serv Res*. 2009 Feb;44(1):182-204.
- 235 Berwick DM. Measuring physicians' quality and performance: adrift on Lake Wobegon. *JAMA*. 2009 Dec 9;302(22):2485-6.
- 236 Spiegelhalter D. Funnel plots for institutional comparison. *Qual Saf Health Care*. 2002 Dec;11(4):390-1.
- 237 Jarman B, Bottle A, Aylin P, Browne M. Monitoring changes in hospital standardised mortality ratios. *BMJ*. 2005 Feb 12;330(7487):329.
- 238 Guthrie B, Love T, Fahey T, Morris A, Sullivan F. Control, compare and communicate: designing control charts to summarise efficiently data from multiple quality indicators. *Qual Saf Health Care*. 2005 Dec;14(6):450-4.
- 239 McLoughlin V, Millar J, Mattke S, Franca M, Jonsson PM, Somekh D, Bates D. Selecting indicators for patient safety at the health system level in OECD countries. *Int J Qual Health Care*. 2006 Sep;18 Suppl 1:14-20.
- 240 Januel JM, Duclos A, Couray-Targe S, Bestion A, Amamra N, Touzet S, Colin C. Développement d'Indicateurs de la Sécurité des Soins (Patient Safety Indicators) à partir des bases de données médico-administratives hospitalières : Evaluation et validation d'une sélection d'indicateur. Rapport d'avancement projet PSI HCL-HAS-DREES, mars 2009 : 26 p.
- 241 Drösler SE, Klazinga NS, Romano PS, Tancredi DJ, Gogorcena Aoiz MA, Hewitt MC, Scobie S, Soop M, Wen E, Quan H, Ghali WA, Mattke S, Kelley E. Application of patient safety indicators internationally: a pilot study among seven countries. *Int J Qual Health Care*. 2009 Aug;21(4):272-8.
- 242 Raleigh VS, Cooper J, Bremner SA, Scobie S. Patient safety indicators for England from hospital administrative data: case-control analysis and comparison with US data. *BMJ*. 2008 Oct 17;337:a1702.
- 243 Fitts P, Posner MI. *Human Performance*. Belmont, CA: Brooks/Cole, 1967.
- 244 Platon. *La République*. Paris, France: Flammarion, 2002.
- 245 Ericsson KA. An expert-performance perspective of research on medical expertise: the study of clinical performance. *Med Educ*. 2007 Dec;41(12):1124-30.
- 246 Choudhry NK, Fletcher RH, Soumerai SB. Systematic review: the relationship between clinical experience and quality of health care. *Ann Intern Med*. 2005 Feb 15;142(4):260-73.
- 247 Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. *Acad Emerg Med*. 2008 Nov;15(11):988-94.

-
- 248 Abernethy B, Poolton JM, Masters RS, Patil NG. Implications of an expertise model for surgical skills training. *ANZ J Surg.* 2008 Dec;78(12):1092-5.
- 249 Ericsson KA, Lehmann AC. *Encyclopedia of Creativity.* San Diego, CA: Academic Press, 1990.
- 250 Grantcharov TP, Reznick RK. Teaching procedural skills. *BMJ.* 2008 May 17;336(7653):1129-31.
- 251 Catchpole K, Mishra A, Handa A, McCulloch P. Teamwork and error in the operating room: analysis of skills and roles. *Ann Surg.* 2008 Apr;247(4):699-706.
- 252 Bates DW, Larizgoitia I, Prasopa-Plaizier N, Jha AK; Research Priority Setting Working Group of the WHO World Alliance for Patient Safety. Global priorities for patient safety research. *BMJ.* 2009 May 14;338:b1775.
- 253 Luft HS, Hunt SS, Maerki SC. The volume-outcome relationship: practice-makes-perfect or selective-referral patterns? *Health Serv Res.* 1987 Jun;22(2):157-82.
- 254 Kozower BD, Stukenborg GJ, Lau CL, Jones DR. Measuring the quality of surgical outcomes in general thoracic surgery: should surgical volume be used to direct patient referrals? *Ann Thorac Surg.* 2008 Nov;86(5):1405-8.
- 255 Or Z, Renaud T. Quel lien entre volume d'activité des hôpitaux et qualité des soins en France ? Institut de Recherche et Documentation en Economie de la Santé. Questions d'économie de la santé 2009 Dec;14:1-6.
- 256 Roger France FH. Case mix use in 25 countries: a migration success but international comparisons failure. *Int J Med Inform.* 2003 Jul;70(2-3):215-9.
- 257 Ordonnance n° 2005-406 du 2 mai 2005 simplifiant le régime juridique des établissements de santé. Titre I, Article 5 (Art. L. 6145-16).
- 258 Wallace JE, Lemaire JB, Ghali WA. Physician wellness: a missing quality indicator. *Lancet.* 2009 Nov 14;374(9702):1714-21.
- 259 Dunstone DC, Reames HR Jr. Physician satisfaction revisited. *Soc Sci Med.* 2001 Mar;52(6):825-37.
- 260 Maslach C, Jackson S, Leiter M. *Maslach Burnout Inventory Manual.* 3rd ed. Palo Alto, CA: Consulting Psychologists Press; 1996.
- 261 Shanafelt TD. Enhancing meaning in work: a prescription for preventing physician burnout and promoting patient-centered care. *JAMA.* 2009 Sep 23;302(12):1338-40.
- 262 Thomas NK. Resident burnout. *JAMA.* 2004 Dec 15;292(23):2880-9.
- 263 Fahrenkopf AM, Sectish TC, Barger LK, Sharek PJ, Lewin D, Chiang VW, Edwards S, Wiedermann BL, Landrigan CP. Rates of medication errors among depressed and burnt out residents: prospective cohort study. *BMJ.* 2008 Mar 1;336(7642):488-91.
- 264 Shanafelt TD, Bradley KA, Wipf JE, Back AL. Burnout and self-reported patient care in an internal medicine residency program. *Ann Intern Med.* 2002 Mar 5;136(5):358-67.
- 265 Landrigan CP, Rothschild JM, Cronin JW, Kaushal R, Burdick E, Katz JT, Lilly CM, Stone PH, Lockley SW, Bates DW, Czeisler CA. Effect of reducing interns' work hours on serious medical errors in intensive care units. *N Engl J Med.* 2004 Oct 28;351(18):1838-48.
- 266 Philibert I. Sleep loss and performance in residents and nonphysicians: a meta-analytic examination. *Sleep.* 2005 Nov 1;28(11):1392-402.
- 267 Dyrbye LN, Thomas MR, Massie FS, Power DV, Eacker A, Harper W, Durning S, Moutier C, Szydlo DW, Novotny PJ, Sloan JA, Shanafelt TD. Burnout and suicidal ideation among U.S. medical students. *Ann Intern Med.* 2008 Sep 2;149(5):334-41.
- 268 Resident duty hours: 1. enhancing sleep, supervision, and safety: Committee on Optimizing Graduate Medical Trainee (Resident) Hours and Work Schedules to Improve Patient Safety. Washington, DC: National Academies Press, 2008.
- 269 Iglehart JK. Revisiting duty-hour limits--IOM recommendations for patient safety and resident education. *N Engl J Med.* 2008 Dec 18;359(25):2633-5.

-
- 270 Kuehn BM. IOM: Shorten residents' work shifts to reduce fatigue, improve patient safety. *JAMA*. 2009 Jan 21;301(3):259-61.
- 271 Ball CG, Kirkpatrick AW, Feliciano DV, Reznick R, McSwain NE. Surgeons and astronauts: so close, yet so far apart. *Can J Surg*. 2008 Aug;51(4):247-50.
- 272 Shanafelt TD, Balch CM, Bechamps GJ, Russell T, Dyrbye L, Satele D, Collicott P, Novotny PJ, Sloan J, Freischlag JA. Burnout and career satisfaction among American surgeons. *Ann Surg*. 2009 Sep;250(3):463-71.
- 273 Shanafelt TD, Balch CM, Bechamps G, Russell T, Dyrbye L, Satele D, Collicott P, Novotny PJ, Sloan J, Freischlag J. Burnout and Medical Errors Among American Surgeons. *Ann Surg*. 2009 Nov 19. [Epub ahead of print]
- 274 Taffinder NJ, McManus IC, Gul Y, Russell RC, Darzi A. Effect of sleep deprivation on surgeons' dexterity on laparoscopy simulator. *Lancet*. 1998 Oct 10;352(9135):1191.
- 275 Rothschild JM, Keohane CA, Rogers S, Gardner R, Lipsitz SR, Salzberg CA, Yu T, Yoon CS, Williams DH, Wien MF, Czeisler CA, Bates DW, Landrigan CP. Risks of complications by attending physicians after performing nighttime procedures. *JAMA*. 2009 Oct 14;302(14):1565-72.
- 276 Rosekind MR, Gander PH, Miller DL, Gregory KB, Smith RM, Weldon KJ, Co EL, McNally KL, Lebacqz JV. Fatigue in operational settings: examples from the aviation environment. *Hum Factors*. 1994 Jun;36(2):327-38.
- 277 Zare MM, Itani KM, Schiffner TL, Henderson WG, Khuri SF. Mortality after nonemergent major surgery performed on Friday versus Monday through Wednesday. *Ann Surg*. 2007 Nov;246(5):866-74.
- 278 Parker JB. The effects of fatigue on physician performance--an underestimated cause of physician impairment and increased patient risk. *Can J Anaesth*. 1987 Sep;34(5):489-95.
- 279 Ramsay MA. Physician fatigue. *Proc (Bayl Univ Med Cent)*. 2000 Apr;13(2):148-50.
- 280 Lingard L, Espin S, Whyte S, Regehr G, Baker GR, Reznick R, Bohnen J, Orser B, Doran D, Grober E. Communication failures in the operating room: an observational classification of recurrent types and effects. *Qual Saf Health Care*. 2004 Oct;13(5):330-4.
- 281 Helmreich RL. On error management: lessons from aviation. *BMJ*. 2000 Mar 18;320(7237):781-5.
- 282 Sexton JB, Thomas EJ, Helmreich RL. Error, stress, and teamwork in medicine and aviation: cross sectional surveys. *BMJ*. 2000 Mar 18;320(7237):745-9.
- 283 McCulloch P, Mishra A, Handa A, Dale T, Hirst G, Catchpole K. The effects of aviation-style non-technical skills training on technical performance and outcome in the operating theatre. *Qual Saf Health Care*. 2009 Apr;18(2):109-15.
- 284 Clergue F. Standardization-communication: two targets for health care safety. *Ann Fr Anesth Reanim*. 2009 May;28(5):423-5.
- 285 Degani, A., & Wiener, E. L. (1990). The human factors of flight deck checklists: The normal checklist. NASA Technical Memorandum #177549. Moffett Field, CA: NASA Ames Research Center.
- 286 Hart EM, Owen H. Errors and omissions in anesthesia: a pilot study using a pilot's checklist. *Anesth Analg*. 2005 Jul;101(1):246-50.
- 287 Lingard L, Regehr G, Orser B, Reznick R, Baker GR, Doran D, Espin S, Bohnen J, Whyte S. Evaluation of a preoperative checklist and team briefing among surgeons, nurses, and anesthesiologists to reduce failures in communication. *Arch Surg*. 2008 Jan;143(1):12-7.
- 288 Nundy S, Mukherjee A, Sexton JB, Pronovost PJ, Knight A, Rowen LC, Duncan M, Syin D, Makary MA. Impact of preoperative briefings on operating room delays: a preliminary report. *Arch Surg*. 2008 Nov;143(11):1068-72.
- 289 Gawande A. *Annals of Medicine : The Checklist*. The New Yorker. 2007 Dec.

-
- 290 Hales BM, Pronovost PJ. The checklist--a tool for error management and performance improvement. *J Crit Care*. 2006 Sep;21(3):231-5.
 - 291 Weiner EL., Kanki BG, Helmreich. RL. *Cockpit Resource Management*. San Diego: Academic Press, 1993.
 - 292 Hamman WR. The complexity of team training: what we have learned from aviation and its applications to medicine. *Qual Saf Health Care*. 2004 Oct;13 Suppl 1:i72-9.
 - 293 Tamuz M, Harrison MI. Improving patient safety in hospitals: Contributions of high-reliability theory and normal accident theory. *Health Serv Res*. 2006 Aug;41(4 Pt 2):1654-76.
 - 294 Singh H, Petersen LA, Thomas EJ. Understanding diagnostic errors in medicine: a lesson from aviation. *Qual Saf Health Care*. 2006 Jun;15(3):159-64.
 - 295 Sockeel P, Chatelain E, Massoure MP, David P, Chapellier X, Buffat S. Surgeons can learn from pilots: human factors in surgery. *J Chir (Paris)*. 2009 Jun;146(3):250-5.
 - 296 Makary MA, Mukherjee A, Sexton JB, Syin D, Goodrich E, Hartmann E, Rowen L, Behrens DC, Marohn M, Pronovost PJ. Operating room briefings and wrong-site surgery. *J Am Coll Surg*. 2007 Feb;204(2):236-43.
 - 297 Lingard L, Espin S, Rubin B, Whyte S, Colmenares M, Baker GR, Doran D, Grober E, Orser B, Bohnen J, Reznick R. Getting teams to talk: development and pilot implementation of a checklist to promote interprofessional communication in the OR. *Qual Saf Health Care*. 2005 Oct;14(5):340-6.
 - 298 Haller G, Garnerin P, Morales MA, Pfister R, Berner M, Irion O, Clergue F, Kern C. Effect of crew resource management training in a multidisciplinary obstetrical setting. *Int J Qual Health Care*. 2008 Aug;20(4):254-63.
 - 299 Davenport DL, Henderson WG, Mosca CL, Khuri SF, Mentzer RM Jr. Risk-adjusted morbidity in teaching hospitals correlates with reported levels of communication and collaboration on surgical teams but not with scale measures of teamwork climate, safety climate, or working conditions. *J Am Coll Surg*. 2007 Dec;205(6):778-84.
 - 300 Gaba DM, Maxwell M, DeAnda A. Anesthetic mishaps: breaking the chain of accident evolution. *Anesthesiology*. 1987 May;66(5):670-6.
 - 301 Helmreich RL, Schaefer HG. *Human Error in Medicine : Team Performance in the Operating Room*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1994.
 - 302 Leonard M, Graham S, Bonacum D. The human factor: the critical importance of effective teamwork and communication in providing safe care. *Qual Saf Health Care*. 2004 Oct;13 Suppl 1:i85-90.
 - 303 McGreevy J, Otten T, Poggi M, Robinson C, Castaneda D, Wade P. The challenge of changing roles and improving surgical care now: Crew Resource Management approach. *Am Surg*. 2006 Nov;72(11):1082-7.
 - 304 Awad SS, Fagan SP, Bellows C, Albo D, Green-Rashad B, De la Garza M, Berger DH. Bridging the communication gap in the operating room with medical team training. *Am J Surg*. 2005 Nov;190(5):770-4.
 - 305 Salas E, Wilson KA, Burke CS, Wightman DC. Does crew resource management training work? An update, an extension, and some critical needs. *Hum Factors*. 2006 Summer;48(2):392-412.

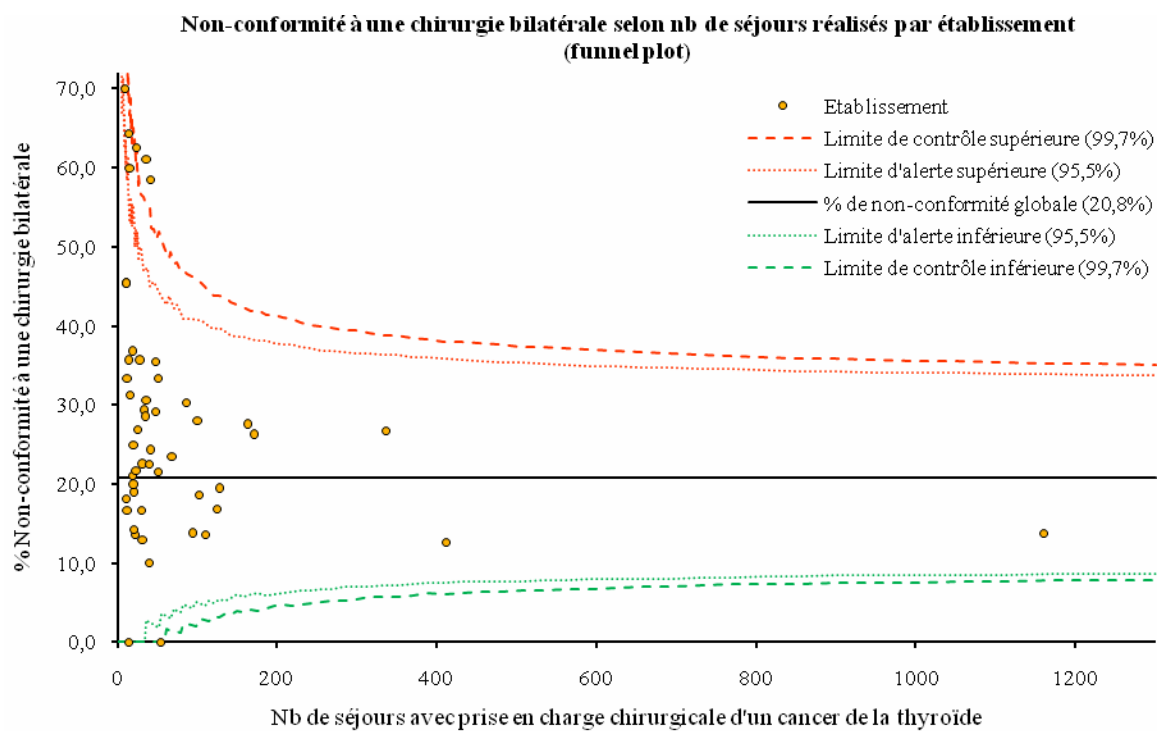
10. Annexes

10.1. Lien entre volume d'activité et indication opératoire (*article 5*)

De nombreuses études ont cherché à quantifier le lien existant entre le volume d'activité et le résultat d'une prise en charge, que ce soit à l'échelle du chirurgien ou d'un établissement (*volume-outcome relationship*). En chirurgie de la thyroïde, il a été montré que la réalisation d'un nombre élevé d'opérations pouvait réduire le risque de complications opératoires. Toutefois, peu d'études ont cherché à étudier le lien entre volume d'activité et qualité de la pratique chirurgicale (*volume-process relationship*). Pour le cancer différencié de la thyroïde, il est recommandé de procéder d'emblée à une chirurgie bilatérale en cas de tumeur présentant une taille supérieure à 1 centimètre.

Nous avons utilisé la base PMSI de Rhône-Alpes dans le but de quantifier le lien existant entre le volume d'opérations thyroïdiennes effectuées par les établissements de la région (toutes pathologies confondues) et leur adhésion à une indication chirurgicale bilatérale dans le cas du cancer. De 1999 à 2004, 20 140 chirurgies thyroïdiennes avaient été réalisées dont 4 006 pour prise en charge d'un cancer. Le taux de non-conformité globale à la recommandation était de 21,1% et le risque de non-conformité à une chirurgie bilatérale d'emblée était significativement supérieur en cas de prise en charge dans un établissement réalisant moins de 100 opérations par an.

Ces résultats étaient en faveur d'une orientation des patients atteints de cancer de la thyroïde dans les centres réalisant un nombre élevé de thyroïdectomies. Toutefois, il convenait de les interpréter prudemment. Un carte de contrôle transversale (*funnel plot*) a été construite afin de représenter la performance individuelle des 51 établissements ayant réalisé au moins 10 séjours pour prise en charge d'un cancer thyroïdien de 1999 à 2004 (figure 17, non présentée dans l'article). Sa lecture révélait que seuls 6 établissements s'écartaient significativement de la moyenne jugée acceptable de non-conformité à une chirurgie bilatérale : 2 établissements se situaient en dehors de la limite de contrôle supérieure et 4 en dehors de celle d'alerte.



Source: Bases PMSI Rhône-Alpes 1999-2004

Figure 17. Carte de contrôle transversale représentant la distribution des établissements en fonction de leur volume de séjours pour prise en charge de cancers thyroïdiens et de leur proportion de non-conformité à la recommandation

Jean-Christophe Lifante, Antoine Duclos, Sandrine Couray-Targe, Cyrille Colin, Jean-Louis Peix, Anne-Marie Schott.

Hospital volume influences the choice of operation for thyroid cancer.

British Journal of Surgery 2009;96:1284-1288.

Le cinquième article publié et reproduit ci-après présente une évaluation du lien existant entre le volume d'opérations sur la thyroïde réalisées par les équipes chirurgicales de Rhône-Alpes et leur respect d'une chirurgie bilatérale d'emblée dans le cancer thyroïdien. Il apporte des éléments de preuve en faveur de la prise en charge du cancer de la thyroïde par des équipes à haut volume d'activité.

Hospital volume influences the choice of operation for thyroid cancer

J. C. Lifante^{1,2}, A. Duclos^{2,3}, S. Couray-Targe^{2,3}, C. Colin^{2,3}, J. L. Peix^{1,2} and A. M. Schott^{2,3}

¹Department of General and Endocrine Surgery, Pôle Chirurgie, Centre Hospitalo Universitaire Lyon-Sud, ²Université Claude Bernard Lyon 1, and ³Hospices Civils de Lyon, Pôle Information Médicale Évaluation Recherche, Lyon, France

Correspondence to: Dr J. C. Lifante, Department of General and Endocrine Surgery, Centre Hospitalo Universitaire Lyon-Sud, Rue du Grand Revoyet, 69495 Pierre Bénite, France

(e-mail: jean-christophe.lifante@chu-lyon.fr)

Background: Many authors advocate total or near-total thyroidectomy for thyroid carcinoma. This study examined the relationship between hospital volume of thyroidectomies and choice of bilateral thyroidectomy for thyroid carcinoma.

Methods: Data were extracted from the administrative databases of all hospital discharge abstracts in the Rhône-Alpes area of France. The study population included inpatient stays from 1999 to 2004 with a diagnosis of thyroid disease (benign or malignant) and a procedural code for thyroid surgery. Multivariable logistic regression analyses were performed to determine factors associated with the extent of surgery (unilateral *versus* bilateral) for thyroid carcinoma.

Results: A total of 20 140 thyroidectomies were identified, including 4006 procedures for cancer. Compared with hospitals performing a high volume of procedures for all thyroid diseases (at least 100 annually), the risk of a unilateral procedure for thyroid cancer increased by 2.46 (95 per cent confidence interval 1.63 to 3.71) in low-volume hospitals (fewer than ten operations per year) and by 1.56 (1.27 to 1.92) in medium-volume centres (ten to 99 per year).

Conclusion: There is a significant relationship between hospital volume and the decision to perform bilateral surgery for thyroid carcinoma. Thyroid cancer surgery should be performed by experienced surgical teams in high-volume centres.

Paper accepted 4 June 2009

Published online in Wiley InterScience (www.bjs.co.uk). DOI: 10.1002/bjs.6741

Introduction

Determining how the volume of a specific operation influences practices in a hospital is a complex process. Total thyroidectomy exposes the patient to the risk of two specific problems: hypoparathyroidism and bilateral recurrent nerve palsy. However, these complications are rare when operations are performed by experienced endocrine surgeons¹. Some authors have reported an association between the choice of total thyroidectomy for multinodular goitre and the surgeon's volume of thyroid procedures. Little has been published on the volume–procedure association in thyroid cancer surgery¹. Currently, total thyroidectomy appears to be the 'gold standard' for well differentiated thyroid carcinomas larger than 1 cm^{2–5}. The aim of this study was to quantify the association between hospital volume of operations for all thyroid disease (benign

and malignant pathology) and the choice of total thyroidectomy for thyroid carcinoma.

Methods

The study was a retrospective analysis of hospital discharge data from all public and private acute-care facilities in the Rhône-Alpes area of France. These facilities are responsible for the care of nearly 6 million people. All patients who underwent thyroid surgery from 1999 to 2004 were considered for inclusion⁶.

Data source

Data were extracted from the French analogue of the Diagnosis Related Group Program, the Programme de Médicalisation des Systèmes d'Information (PMSI). The

PMSI database includes standard discharge abstracts that contain compulsory information about the patient (sex and age), diagnoses (using the International Classification of Diseases (ICD) 10), and the care received (using procedural codes of the French medical procedure catalogue and the Obis version of the common classification of medical procedures) for each hospital stay^{7,8}.

Identification of patients who had thyroid surgery

The included cases were inpatient stays with a diagnosis of thyroid disease or disordered thyroid gland (ICD-10 code E00–E07), malignant neoplasm (-C73) or neoplasm of uncertain behaviour (-D44.0) accompanied by a procedural code for thyroid surgery. This type of algorithm for case identification has been described previously⁹. The study population was separated into two groups based on the extent of surgery performed. One group included stays in which a unilateral thyroid procedure (such as lobectomy) was performed, whereas the other group included stays with a bilateral procedure (near-total thyroidectomy, total thyroidectomy, extended thyroidectomy or secondary completion of thyroidectomy). The cancer-related diagnoses that appeared in the PMSI database were always based on the preoperative findings. In France, only thyroid nodules larger than 1 cm are investigated by fine-needle aspiration before surgery. Thus, only thyroid cancers larger than 1 cm were expected to be identified in the study population.

Hospital procedure volume

Hospital procedure volume was modelled as a categorical variable. For each hospital, the total number of stays for operations relating only to cancer and those relating to all thyroid disease were quantified. The hospitals were categorized according to their annual volume of surgical procedures as follows: fewer than ten, ten to 99 or at least 100 operations per year. Other study variables included patient sex and age, hospital category (public, private for profit, or teaching) and year of admission to hospital.

Statistical analysis

The total numbers of operations per year performed by each hospital both for all thyroid diseases and cancer alone were compared using Spearman's rank correlation test. Univariable statistics were obtained using the χ^2 test to compare the frequencies of categorical variables between hospital admissions for unilateral *versus* bilateral surgery. Factors associated with the extent of the surgery

Table 1 Distribution of hospital stays for unilateral and bilateral thyroid cancer surgery per year

	All procedures	Unilateral	Bilateral
1999	509	127 (25.0)	382 (75.0)
2000	580	115 (19.8)	465 (80.2)
2001	762	182 (23.9)	580 (76.1)
2002	706	139 (19.7)	567 (80.3)
2003	716	149 (20.8)	567 (79.2)
2004	733	132 (18.0)	601 (82.0)
All years	4006	844 (21.1)	3162 (78.9)

Values in parentheses are percentages.

Table 2 Annual distribution of hospitals according to hospital volume category

	Thyroidectomies per year		
	< 10	10–99	≥ 100
1999	9 (16)	37 (67)	9 (16)
2000	7 (13)	39 (72)	8 (15)
2001	10 (17)	40 (68)	9 (15)
2002	14 (23)	34 (57)	12 (20)
2003	11 (19)	34 (60)	12 (21)
2004	11 (20)	33 (60)	11 (20)
All years	62 (18.2)	217 (63.8)	61 (17.9)

Values in parentheses are percentages.

(unilateral *versus* bilateral surgery as the dependent variable) were assessed by crude and adjusted odd ratios (ORs) derived from univariable and multivariable logistic regression models. The variables considered for adjustment were sex, age, hospital category and year of hospital admission. The final multivariable models were based on co-variables identified by clinical judgement in conjunction with automated variable selection methods (the backward stepwise method according to selection criteria with the likelihood ratio method, where $P_{in} = 0.100$ and $P_{out} = 0.150$). In addition, the co-variables were assessed for potential interactions. Analyses were performed successively by first considering the hospital volume for all thyroid diseases (including cancer) and then for cancer alone. All analyses were conducted using SAS[®] software, release 9.13 (SAS Institute, Cary, North Carolina, USA). All *P* values were two sided, and 0.050 was used as the threshold for statistical significance.

Results

A total of 20140 hospital stays for thyroid operations were identified between 1999 and 2004. Among these, 4006 procedures involved patients with cancer of the

Table 3 Univariable analysis of surgical procedures for thyroid cancer according to patient and hospital characteristics

	Surgical procedure			<i>P</i> ‡
	All (<i>n</i> = 4006)	Unilateral (<i>n</i> = 844)	Bilateral (<i>n</i> = 3162)	
Type of surgery				
Lobectomy	844 (21.1)	844 (100)	0 (0)	
Near-total thyroidectomy	184 (4.6)	0 (0)	184 (5.8)	
Total thyroidectomy	2339 (58.4)	0 (0)	2339 (74.0)	
Extended thyroidectomy	126 (3.1)	0 (0)	126 (4.0)	
Completion of thyroidectomy	513 (12.8)	0 (0)	513 (16.2)	
Sex				0.797
F	3048 (76.1)	645 (76.4)	2403 (76.0)	
M	958 (23.9)	199 (23.6)	759 (24.0)	
Age (years)				0.001
< 45	1568 (39.1)	371 (44.0)	1197 (37.9)	
≥ 45	2438 (60.9)	473 (56.0)	1965 (62.1)	
Healthcare sector				< 0.001
Public hospital	708 (17.7)	161 (19.1)	547 (17.3)	
Private hospital	1173 (29.3)	332 (39.3)	841 (26.6)	
Teaching hospital	2125 (53.0)	351 (41.6)	1774 (56.1)	
Hospital volume for thyroid disease*				< 0.001
< 10	117 (2.9)	44 (5.2)	73 (2.3)	
10–99	1123 (28.0)	323 (38.3)	800 (25.3)	
≥ 100	2766 (69.0)	477 (56.5)	2289 (72.4)	
Hospital volume for thyroid cancer†				< 0.001
< 10	901 (22.5)	272 (32.2)	629 (19.9)	
10–99	1611 (40.2)	372 (44.1)	1239 (39.2)	
≥ 100	1494 (37.3)	200 (23.7)	1294 (40.9)	

Values in parentheses are percentages. *Hospital annual volume of stays with a surgical procedure for thyroid disease, including cancer; †hospital annual volume of stays with a surgical procedure for thyroid cancer. ‡ χ^2 test.

thyroid gland at a total of 77 hospitals (35 per cent public, 55 per cent private and 10 per cent teaching hospitals). The median age of the patients was 49 (range 2–94) years. Seven patients died before hospital discharge (0.2 per cent). The mean(s.d.) length of stay was 4.4(4.9) days. The number of stays for thyroid cancer surgery varied over time (*Table 1*). Each year, 54–60 hospitals performed operations for thyroid cancer. With regard to all thyroid diseases, *Table 2* shows the number of centres performing fewer than ten procedures per year, between 10 and 99 per year, and more than 100 per year. The annual thyroid-specific volume at each hospital was strongly correlated with the annual thyroid cancer-specific volume ($r_s = 0.82$, $P < 0.001$).

Factors associated with unilateral surgery

Of the 4006 procedures performed for thyroid cancer, 844 (21.1 per cent) were unilateral and 3162 (78.9 per cent) were bilateral. The univariable statistics showed that patients who had a unilateral procedure were more likely to be under 45 years old, be treated at a private hospital, and receive care at a hospital with an annual procedure volume of less than 100 per year (*Table 3*). Stays in low-volume

hospitals were independently associated with a higher risk of unilateral surgery for thyroid carcinoma after adjustment for age, hospital category and year of hospital stay ($P < 0.001$) (*Table 4*). Compared with hospitals with a high volume of procedures for all thyroid diseases (at least 100 per year), the risks of a unilateral procedure for thyroid cancer increased by 2.46 (95 per cent confidence interval 1.63 to 3.71) and 1.56 (1.27 to 1.92) in low-volume (less than ten per year) and medium-volume (ten to 99 per year) hospitals respectively.

Discussion

The present study showed a strong association between bilateral surgery for thyroid cancer and hospital thyroidectomy volume. This finding highlights the fact that the choice between unilateral and bilateral procedures for thyroid carcinoma depends on the volume of procedures performed for all thyroid diseases (benign and malignant).

Total or near-total thyroidectomy is often considered for thyroid cancers larger than 1 cm owing to the risk of recurrence^{4,5}. In general, the practice guidelines for thyroid cancer advocate total thyroidectomy for

Table 4 Logistic regression analysis of factors associated with a unilateral procedure for thyroid cancer surgery

	Unilateral <i>versus</i> bilateral				
	Crude OR§	Adjusted OR¶	P	Adjusted OR#	P
Age (years)					
< 45	1.29 (1.10, 1.50)	1.35 (1.16, 1.58)	< 0.001	1.37 (1.17, 1.61)	< 0.001
≥ 45*	1.00	1.00		1.00	
Healthcare sector			< 0.001		0.027
Public hospital	1.49 (1.21, 1.83)	1.20 (0.95, 1.51)		0.83 (0.64, 1.08)	
Private hospital	1.99 (1.68, 2.37)	1.54 (1.24, 1.92)		1.13 (0.88, 1.43)	
Teaching hospital*	1.00	1.00		1.00	
Hospital volume for thyroid disease†			< 0.001		
< 10	2.89 (1.96, 4.26)	2.46 (1.63, 3.71)		—	
10–99	1.94 (1.65, 2.28)	1.56 (1.27, 1.92)		—	
≥ 100*	1.00	1.00		—	
Hospital volume for thyroid cancer‡					< 0.001
< 10	2.80 (2.28, 3.44)	—		2.85 (2.12, 3.85)	
10–99	1.94 (1.61, 2.35)	—		2.03 (1.60, 2.58)	
≥ 100*	1.00	—		1.00	
Year of hospital admission			0.003		0.003
1999	1.51 (1.15, 1.99)	1.60 (1.21, 2.12)		1.57 (1.18, 2.08)	
2000	1.13 (0.85, 1.49)	1.16 (0.87, 1.53)		1.16 (0.87, 1.54)	
2001	1.43 (1.11, 1.84)	1.57 (1.21, 2.02)		1.58 (1.22, 2.05)	
2002	1.12 (0.86, 1.45)	1.19 (0.91, 1.56)		1.18 (0.90, 1.55)	
2003	1.20 (0.92, 1.55)	1.22 (0.94, 1.60)		1.22 (0.93, 1.59)	
2004*	1.00	1.00		1.00	

Values in parentheses are 95 per cent confidence intervals. *Reference group. †Hospital annual volume of stays with a surgical procedure for thyroid disease, including cancer; ‡hospital annual volume of stays with a surgical procedure for thyroid cancer. §Crude odds ratio (OR) from logistic regression analysis with unilateral procedure as the dependent variable and without adjustment for co-variables. ¶Adjusted OR from logistic regression analysis with unilateral procedure as the dependent variable, and age, healthcare sector, hospital volume for thyroid disease and year as co-variables for adjustment (4006 stays included, with 844 unilateral procedures *versus* 3162 bilateral procedures as reference). #Adjusted OR from logistic regression analysis with unilateral procedure as the dependent variable, and age, healthcare sector, hospital volume for thyroid cancer and year as co-variables for adjustment (4006 stays included, with 844 unilateral procedures *versus* 3162 bilateral procedures as reference).

carcinomas larger than 1 cm^{10–13}. Total thyroidectomy also allows radioiodine ablation and better surveillance using thyroglobulin assays. Thus, performing bilateral thyroidectomy for cancers larger than 1 cm might be considered an indicator of surgical quality.

Sosa and co-workers¹⁴ have already established a relationship between surgeon thyroidectomy volume and outcome, based on postoperative complications including recurrent nerve injury and hypoparathyroidism. Quality measurement based on outcome indicators from diagnosis-related group administrative databases might be biased in thyroid surgery. In addition to the variability in data collection among providers, it is not possible to distinguish between patient complications and comorbidities^{14,15}. Moreover, the specific complications coded in these databases are not reliable because they depend on clinical and technical tests used for diagnosis. For example, it is well known that there is variability in the detection of recurrent laryngeal nerve palsy and hypocalcaemia^{16,17}. Accordingly, a process indicator appears to be more appropriate than

outcome measures for assessing the quality of thyroid surgery.

High-volume surgeons are associated with better patient outcomes, shorter length of hospital stay and lower inpatient costs in paediatric endocrine surgery¹⁸ as well as thyroid surgery in the elderly¹⁹. These issues relate mainly to highly specialized surgery where practice evolution requires ongoing updates of literature advances²⁰. The present study provides an additional argument for the management of thyroid carcinoma in specialized centres. Reoperations owing to errors of judgement are associated with low volume²¹. Lack of experience may lead to inadequate surgery for thyroid carcinoma owing to insufficient knowledge regarding the current standards of care for specialized diseases²¹.

The present study has several potential limitations. The lack of information regarding tumour size in the PMSI database meant that it was not possible to differentiate tumours smaller than 1 cm from larger ones. However, this did not influence the results. Indeed, the coded diagnoses in the database were always based on preoperative findings,

and only nodules larger than 1 cm are investigated before surgery in France. Because microcarcinomas are usually found during the pathological examination of thyroid specimens removed for another disease, they would not contribute to the diagnoses coded in the database. In addition, it is unclear whether the present findings could be extended to other geographical areas; they may only reflect the surgical practices in the Rhône-Alpes area. However, a major strength of this study lies in its population-based design. Further worldwide studies are needed to corroborate the potential link between the volume of procedures and the quality of the process in thyroid surgery.

Acknowledgements

The authors declare no conflict of interest.

References

- Thomusch O, Machens A, Sekulla C, Ukkat J, Lippert H, Gastinger I *et al.* Multivariate analysis of risk factors for postoperative complications in benign goiter surgery: prospective multicenter study in Germany. *World J Surg* 2000; **24**: 1335–1441.
- Pacini F, Schlumberger M, Harmer C, Berg GG, Cohen O, Duntas L *et al.* Post-surgical use of radioiodine (^{131}I) in patients with papillary and follicular thyroid cancer and the issue of remnant ablation: a consensus report. *Eur J Endocrinol* 2005; **153**: 651–659.
- Sosa JA, Udelsman R. Total thyroidectomy for differentiated thyroid cancer. *J Surg Oncol* 2006; **94**: 701–707.
- Hay ID, Grant CS, Bergstralh EJ, Thompson GB, van Heerden JA, Goellner JR. Unilateral total lobectomy: is it sufficient surgical treatment for patients with AMES low-risk papillary thyroid carcinoma? *Surgery* 1998; **124**: 958–964.
- Mazzaferri EL, Kloos RT. Clinical review 128: Current approaches to primary therapy for papillary and follicular thyroid cancer. *J Clin Endocrinol Metab* 2001; **86**: 1447–1463.
- Institut National de la Statistique et des Études (INSEE). *Résultats Régionaux des Enquêtes de Recensement de 2004 à 2007*. http://www.insee.fr/fr/fr/recensement/nouv_recens/resultats-regionaux.htm [accessed 9 January 2009].
- French Technical Hospitalisation Agency (ATIH). *PMSI MCO Presentation: Understanding the Principles and Background of this PMSI Domain*. <http://www.atih.sante.fr/en/index.php?id=0002300005FF> [accessed 9 January 2009].
- World Health Organization. *International Statistical Classification of Diseases and Related Health Problems*, 10th revision. <http://www.who.int/classifications/apps/icd/icd10online/> [accessed 9 January 2009].
- Couris CM, Duclos A, Rabilloud M, Couray-Targe S, Ecochard R, Delmas PD *et al.* A seventy percent overestimation of the burden of hip fractures in women aged 85 and over. *Bone* 2007; **41**: 896–900.
- Bilimoria KY, Bentrem DJ, Ko CY, Stewart AK, Winchester DP, Talamonti MS *et al.* Extent of surgery affects survival for papillary thyroid cancer. *Ann Surg* 2007; **246**: 375–381.
- British Association of Endocrine Surgeons (BAETS). Surgery for differentiated thyroid cancer. In *Report of the Thyroid Cancer Guidelines Update Group: Royal College of Physicians*, Perros P (ed.). Royal College of Physicians: London, 2007; 12–17.
- Cooper DS, Doherty GM, Haugen BR, Kloos RT, Lee SL, Mandel S *et al.* Management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid* 2006; **16**: 109–142.
- National Comprehensive Cancer Network (NCCN). *Clinical Practice Guidelines in Oncology: Thyroid Carcinoma V.1.2008*. 2008; http://www.nccn.org/professionals/physician_gls/PDF/thyroid.pdf [accessed 29 January 2009].
- Sosa JA, Bowman HM, Tielsch JM, Powe NR, Gordon TA, Udelsman R. The importance of surgeon experience for clinical and economic outcomes from thyroidectomy. *Ann Surg* 1998; **228**: 320–330.
- Roos LL, Stranc L, James RC, Li J. Complications, comorbidities, and mortality: improving classification and prediction. *Health Serv Res* 1997; **32**: 229–238.
- Trésallet C, Chigot J, Menegaux F. How to prevent recurrent nerve palsy during thyroid surgery? *Ann Chir* 2006; **131**: 149–153.
- Proye C, Patoir A, Trincaretto F, Darras J, Lagache G. [Advantages of systematic recurrent laryngeal nerve resection during thyroid surgery. Results of 422 thyroidectomies employing this procedure.] *J Chir (Paris)* 1980; **117**: 155–160.
- Tuggle TC, Roman SA, Wang TS, Boudourakis L, Thomas DC, Udelsman R *et al.* Pediatric endocrine surgery: who is operating on our children? *Surgery* 2008; **144**: 869–877.
- Sosa JA, Mehta JP, Wang TS, Boudourakis L, Roman SA. A population-based study of outcomes from thyroidectomy in aging Americans: at what cost? *J Am Coll Surg* 2008; **206**: 1097–1105.
- Hogan AM, Winter DC. Does practice make perfect? *Ann Surg Oncol* 2008; **15**: 1267–1270.
- Mitchell J, Milas M, Barbosa G, Sutton J, Berber E, Siperstein A. Avoidable reoperations for thyroid and parathyroid surgery: effect of hospital volume. *Surgery* 2008; **144**: 899–906.

10.2. Fiche de codage

Évaluation de la chirurgie thyroïdienne : Fiche de codage PMSI

Critère d'inclusion : patient hospitalisé pour bénéficier d'une intervention sur la thyroïde réalisée par l'un des chirurgiens du service.

Critères d'exclusion : erreur de codage ou donnée manquante, patient présentant une hypoparathyroïdie ou une paralysie récurrentielle antérieure à l'intervention.

Règles de codage : Cette grille doit être complétée à l'aide du dossier clinique du patient (compte-rendu opératoire, biologie, laryngoscopie). Il est nécessaire de coder chaque variable en entourant le numéro ou en renseignant le champ correspondants.

Règles de saisie : À la sortie du patient, lors du recueil de son séjour dans SIM Web100T, cliquer sur le volet 'PMSI' puis renseigner le tableau des 'Codes internes' à l'aide de cette fiche de codage préalablement complétée. Pour chacune des variables, il est nécessaire de réinscrire le N° correspondant (une lettre majuscule) et son Code.

Nature de l'acte réalisé

Code = **[N]** 1 : Lobo-isthmectomie
2 : Thyroïdectomie totale ou subtotale
3 : Réintervention thyroïdienne
4 : Thyroïdectomie pour cancer avec au minimum un curage central
5 : Autres interventions sur la thyroïde
9 : Donnée non disponible

Paralysie récurrentielle postopératoire

Code = **[P]** 1 : Unilatérale ou Bilatérale
2 : Absence de paralysie
9 : Donnée non disponible (ou PR non confirmée)

Calcémie postopératoire la plus basse

Code = **[C]** 1 : < 2 mmol/l
2 : ≥ 2 mmol/l
9 : Donnée non disponible (ou absence de dosage post-opératoire)

Identité du chirurgien responsable de l'intervention (depuis mai 2008)

Code = **[I]** Initiales du chirurgien (Prénom + Nom) : ____

Exemple : JLP, JCL, LM, SO, DV, etc.

Monitoring du nerf par électrostimulation lors de l'intervention (depuis mai 2008)

Code = **[M]** 1 : Oui
2 : Non
9 : Donnée non disponible

ATTENTION :

**Cette grille doit être conservée dans le dossier du patient.
L'exploitation des données recueillies nécessite qu'elles soient codées avec exactitude et de manière standardisée, dans le respect des règles de la grille de codage.**

« Un demi siècle s'est écoulé depuis la parution du livre de Shewhart. Un demi-siècle s'écoulera encore avant que la pensée de Shewhart soit entièrement connue dans l'enseignement, la science, et l'industrie. »

W. Edwards Deming, 1981.^{VIII}

^{VIII} Shewhart WA, Deming WE. Statistical Method from the Viewpoint of Quality Control. 2nd ed. New York: Dover, 1986.

TITRE : Sécurité du patient en chirurgie thyroïdienne : intérêt du suivi des complications par cartes de contrôle.

RÉSUMÉ : À l'hôpital, le bloc opératoire représente un environnement à risque pour le patient. La réalisation d'une chirurgie thyroïdienne demeure une tâche complexe dont le résultat dépend de la combinaison d'une multitude de facteurs qui demeurent mal connus. L'application des méthodes de contrôle qualité développées dans l'industrie peut être utile à leur découverte. Pendant quatre années, nous avons suivi la performance collective et individuelle d'une équipe chirurgicale à l'aide de cartes de contrôle. Des cartes de Shewhart et des cartes CUSUM ont été conçues pour analyser les variations des complications de la thyroïdectomie au cours du temps. Leur mise en place s'est accompagnée d'une réduction immédiate des hypocalcémies postopératoires, puis d'une stabilisation progressive des paralysies récurrentielles. La suractivité chirurgicale a été identifiée comme cause à l'origine d'une dégradation momentanée de la performance de l'équipe. Cette expérience de terrain a montré l'applicabilité des cartes de contrôle à la chirurgie thyroïdienne ainsi que leur intérêt pour aider les chirurgiens à interpréter la variabilité de leurs résultats et à maîtriser les facteurs y concourant. Cependant, seules des actions concrètes destinées à améliorer la qualité de la prise en charge peuvent réduire la survenue de complications opératoires. Leur conception implique de mieux comprendre les facteurs influençant la performance du chirurgien et le travail en équipe au bloc opératoire pour garantir la sécurité du patient.

MOTS-CLÉS : *Sécurité du patient, Qualité des soins, Chirurgie de la thyroïde, Carte de contrôle.*

TITLE: Patient safety in thyroid surgery: value of complications monitoring using control charts.

ABSTRACT: The operating room represents a high risk environment for the patient. Thyroid surgery remains a complex task whose outcome depends on the combination of poorly understood factors. Implementation of industrial quality control methods can be helpful in discovering them. For four years, we have monitored the individual and collective performance of a surgical team using control charts. Shewhart and CUSUM charts have been designed for analyzing the variations of thyroidectomy related complications over time. Their introduction immediately led to a reduction of postoperative hypocalcaemia, and then to a gradual stabilization of recurrent nerve palsy. The surgical overactivity was identified as a root cause of a temporary worsening of the team performance. This field study has shown the applicability of control charts in thyroid surgery and their value to assist surgeons in interpreting and controlling their results' variability. However, only concrete actions aiming at improving the quality of care may reduce the incidence of surgical complications. Designing these actions involves understanding the factors that influence the surgeon's performance and teamwork within operating room to ensure the patient safety.

KEYWORDS: *Patient safety, Quality of care, Thyroid surgery, Control chart.*

DISCIPLINE : Épidémiologie & Sciences de Gestion

INTITULE ET ADRESSE DU LABORATOIRE :

Université de Lyon, Équipe d'Accueil Santé-Individu-Société 4129, Hôtel-Dieu, Place de l'Hôpital, Lyon, F-69002 France.

Hospices Civils de Lyon, Pôle Information Médicale Évaluation Recherche, 162 avenue Lacassagne, Lyon, F-69003, France